REMARKS

Reconsideration of the above-identified application in view of the amendments above and the remarks following is respectfully requested.

Claims 1-16 are in this case. Claims 1-16 have been rejected. Claims 3, 4, 6-10 and 12-14 have now been canceled. Claims 1, 2, 5, 11, 15 and 16 have now been amended.

35 U.S.C. § 112, First Paragraph, Rejections

The Examiner has rejected claims 1, 11 and 16 and dependents under 35 U.S.C. §112, first paragraph, as failing to comply with the written description requirement. The Examiner's rejections are respectfully traversed. Claims 1, 11 and 16 have now been amended.

The Examiner states that the claims contain the phrase " ... wherein the ripe fruit has lost at least 30% of its red ripe fruit water content" and that there is no basis for this phrase in the specification.

Applicant has elected to remove this phrase from the claims thereby overcoming the Examiner's rejections with respect to this phrase.

The Examiner also repeats the rejection of claims 1-16 as failing to comply with the written description requirement as set forth in the Office Action dated June 9, 2004.

The instant application provides guidelines for utilizing at least one wild tomato species (L. hirsutum) while pointing out the suitability of any Lycopersicon species. Since few other crops are blessed with such extensive collections of wild forms and their derivatives and since various wild members of the Lycopersicon genus are well known, one of ordinary skill in the art privileged to the teachings of the present invention would be well aware of the various species covered by the phrase " Lycopersicon spp." recited in the instant application.

Prior art studies investigating commercially important traits in tomato have shown that most if not all of the wild Lycopersicon species studied are suitable

donors for such traits, as is the case with the trait described in U.S. Patent Nos. 5,434,344, 5,817,913, and 6,720,485.

Thus, it is Applicant's strong opinion that the instant application provides the written description support necessary for one of ordinary skill in the art to make and use the present invention as claimed.

35 U.S.C. §112, First Paragraph, Rejections

The Examiner has rejected claims 1-16 under 35 U.S.C. §112, first paragraph, as containing subject matter which was not described in the specification in such a way as to enable one skilled in the relevant art to make and/or use the invention. The Examiner's rejections are respectfully traversed. Claims 3, 4, 6-10 and 12-14 have now been cancelled rendering moot the Examiner's rejections with respect to these claims. Claims 1, 2, 5, 11 15 and 16 have now been amended.

In particular, the Examiner points out that Applicant has not provided evidence that other Lycopersicon species can be crossed with *L. esculentum* to obtain a tomato fruit characterized by the capability of natural dehydration while on the tomato plant.

Applicant contends that actual experimental evidence would not be necessary to convince the ordinary skilled artisan of the suitability of other wild tomato species simply because it is well known that the wild species of Lycopersicon show genetic conservation especially with respect to genetic characteristics related to fruit quality (Frary et al., Theoretical and Applied Genetics, 2004, 108:485-496).

Such conservation is explained by the fact that evolution of the cultivated tomato was driven by domestication which emphasized fruit quality as a marker. As a result, selection of plants occurred particularly with regard to the edible fruit, which served as one of the major foci of selection during the domestication process. Thus, it can be assumed that fruit traits present in wild species which subsequently underwent evolutionary events associated with the domestication process would likely be retained across the wild species. For example, alleles for small fruit size can be found among numerous wild species, while large fruit size is a trait associated with

the more recently evolved cultivated species L. esculentum.

In 2002 Nesbitt and Tanksley (enclosed herewith) surveyed the various wild Lycopersicon species in order to determine the evolutionary events leading to large fruit in the cultivated tomato; they found that the fw2.2 locus in the various wild species is different from that of cultivated tomato. Furthermore, Frary et al., (*ibid*) also clearly showed that a number of fruit-related and domestication-related traits are conserved among the wild species and distinct from the cultivated L. esculentum. Frary et al. summarized the data from previous studies of QTLs in the Lycopersicon genus and showed that at least 3 wild species have orthologues for the epidermal reticulation (*er*) trait. They also report orthologues between wild species for other fruit quality traits, and note that "The majority (76%) of the putatively conserved loci were identified in three or more populations derived from different tomato species" (pg. 495). Thus, Frary et al. (and others) clearly propose that fruit related traits are highly conserved among wild tomato species.

Similarly, the trait of sucrose accumulation, controlled by the invertase gene locus is present in numerous wild species of Lycopersicon which exhibit loss of expression of the invertase gene. During the evolution of Lycopersicon, expression of the invertase gene was activated causing invertase enzyme activity during fruit development and subsequent hydrolysis of sucrose to the hexose sugars, glucose and fructose (U.S. Patent No. 5,434,344; poster of Dai et al., enclosed). Since all the wild species of the subgenus Eriopersicon of the Lycopersicon genus maintain the allelic status associated with the pre-evolutionary event (i.e. no invertase function), it stands to reason that other fruit related genetic traits associated with fruit quality will also be maintained in their pre-evolutionary state in the wild tomato species.

Additional fruit qualities such as fructose to glucose ratio and starch content are also conserved in wild tomato species, again showing that wild tomato species are highly conserved with respect to fruit traits.

The molecular evolutionary event that occurred during the evolution of the cultivated *L. esculentum*, which led to a fruit cuticle impervious to extreme water loss during the final stages of ripening and post harvest, also indicates that the other wild

species of Lycopersicon are a suitable source for the trait of fruit dehydration.

The present inventor has shown that the molecular event that led to a fruit cuticle impervious to extreme water loss is silencing of the expression of the CWP gene in the developing tomato fruit. CWP (for *c*uticular *w*ater *p*ermeability) encodes for a putative protein and the gene is expressed in the immature fruit skin tissue of all the wild species studied, while cultivated L. esculentum tomato plants studied showed no expression of this gene. Silencing of this gene is necessary for the development of cultivated tomatoes devoid of the trait of dehydration and fruit wrinkling and was apparently selected for during the domestication of L. esculentum. A study conducted by the present inventor following filing of the instant application (see Appendix B) demonstrated that the CWP gene is expressed by several wild species including L. pennellii, L. chmieliewskii, L. peruvianum, L. pimpinellifolium and L. cheesmanii. Only in the L. esculentum cultivars and the primitive forms of L. esculentum var. cerasiforme (small fruited forms of the cultivated L. esculentum) accessions is the gene not expressed.

The absence of CWP gene expression in L. esculentum allows development of a cuticle devoid of microfissures. Further proof to this function of CWP was provided in a study performed following filing of the instant application in which the present inventor conclusively showed that transgenic expression of the CWP gene in a cultivated tomato causes fruit microfissures and dehydration identical to that of the cultivated tomato produced according to the teachings of the present invention.

This study also demonstrated that developing fruit of the wild species L. hirsutum, as well as other wild species (including L. chmiewiliewskii, L. pennellii, L. peruvianum, L. cheesmanii and L. pimpinellifolium) all retain the wild genetic trait of CWP gene expression absent from L. esculentum or L. esculentum var. cerasiforme further supporting the fact that any wild species of the Lycopersicon genus can be a source of the trait of fruit dehydration.

Therefore, in light of the genetic conservation present in wild tomato species especially with respect to fruit traits and the fact that such conservation and thus the suitability of use of any wild tomato species in the method of the present invention has

been confirmed by the inventors as well as other research groups, it is Applicant's strong opinion that the instant application provides the guidelines and support necessary for using the method of amended claim 1 to produce the tomato fruit of amended claims 15 and 16 in a reproducible manner with high expectation of success.

The Examiner further states that although the specification discloses the steps involved in the derivation of the tomato plants of the present invention, it is known in the art that introgression of alleles from one background to another is unpredictable. In fact, the specification states that only 25 of the 350 F2 plants (~7%) produced fruit, and of those 25 F2 plants only 3 (~0.8%) were selected for the desired trait. The Examiner concludes by stating that this does not constitute a readily obtainable, repeatable method, and thus it is necessary for the Applicant to meet the deposit requirement.

The Examiner statement that the present method is unpredictable and not reproducible is erroneous in light of the fact that selection of present plants is facilitated via a highly distinguishable and readily observable phenotype, namely wrinkled fruit skin which is raisin-like in appearance (see Figures 1a-c of Appendix A). Applicant would like to point out that since the present plants possess such a readily discernable phenotype, it would be possible to select plants even in cases where successful introgression occurs in only 1 of 1000 plants, much like blue/white color selection enables identification of one correctly transformed bacteria out of a large population of 1000 colonies.

In addition, data that 3/25 fruiting plants possessed the wrinkling trait in the F2 progeny is irrelevant since only 25/350 F2 plants set fruit as expected from plants generated by interspecific crosses. As with the non-wrinkling fruits, non-fruiting plants are easy to discount. Inheritance data of 3/25 clearly supports the reproducibility and predictability of the present method. The fact that the inventor later showed that the *CWP wrinkling* trait is inherited as a single major gene further supports these results.

Selection of plants in this case is similar to selection of transformed bacteria via blue/white colony selection (LacZ expression). Clearly, the present method leads

to a higher rate of success than most bacterial transformation events and yet, blue/white colony selection is considered a predictable and reproducible method, even if transformants comprise less than 0.1% of plated colonies.

In view of the above arguments, Applicant believes to have overcome the 35 U.S.C. §112, first paragraph, rejections.

35 U.S.C. §102(b) Rejections – Eshed and Zamir

The Examiner has rejected claims 15-16 under 35 U.S.C. § 102(b) as being anticipated by Eshed and Zamir. The Examiner's rejections are respectfully traversed. Claims 15 and 16 have now been amended.

The Examiner states that the Schaffer declaration of December 7, 2004 stated that this reference showed the microfissures and dehydration phenotype claimed by the present invention.

Applicant would like to clarify any confusion caused by the plants generated by Eshed and Zamir.

Eshed and Zamir described a NILs, (near introgression lines) platform and outline the value of such a platform in genetic and breeding research. Although Eshed and Zamir described various introgressions from wild tomato species into *L. esculentum* including the IL4-4 introgression, they did not describe the epidermal reticulation shown in Figure 4c of Appendix A, nor did they describe whole fruit dehydration and wrinkling while fruit were either attached to the vine (Figures 1a-b of Appendix A) or detached therefrom (Figures 5a-c of Appendix A). Simply put, Eshed and Zamir worked on an introgression line capable of fruit wrinkling <u>but they did not identify nor isolate plants</u> having wrinkled fruit skin phenotype simply because they did not grow the plants to a point where they produced dehydrated, wrinkled raisin-like fruit.

Eshed and Zamir demonstrated that introgressions from wild species can contribute to horticultural traits such as Brix, fruit yield and other traits that are measured on a mature fruit harvested at the commercially edible stage, prior to a pre-wrinkling stage, if one exists. Such traits cannot be measured on whole fruits past the

stage of ripening or following a delayed storage period since fruit quality generally degrades as fruit remain on the vine past their optimum development and ripening. Maintaining fruit for the length of period that would show wrinkling in the case of the trait being present would be accompanied by fruit degradation and the inability to carry out the destructive quality measurements normally carried out in the traditional selection schemes. Therefore, Brix, which necessitates destruction of fruit is measured on picked ripe fruit but is not measured on fruit past the fully ripe stage. This explains why Eshed and Zamir did not observe natural fruit dehydration since plants and fruit were not grown past the stage of fruit ripening and therefore the selection procedure described in the invention was not practiced thereby.

In order to further distinguish the present invention as claimed from the teachings of the prior art , claims 15 and 16 have now been amended to recite isolated fruit "characterized by skin wrinkling caused by natural fruit dehydration" (claim 15) or characterized by skin wrinkling and an untreated skin" (claim 16). Both claims now clearly describe subject matter which is neither anticipated nor rendered obvious by Eshed and Zamir.

The Examiner has also rejected claims 15 and 16 under 35 U.S.C. §102(b) as being anticipated by Eshed and Zamir in light of the Schaffer declaration filed with the previous response.

As is argued above, Applicant is of the opinion that Eshed and Zamir could not have identified wrinkled dehydrated fruit simply because such a trait appears post ripening, a stage which has no use in the experiments conducted by Eshed and Zamir.

35 U.S.C. §102(b) Rejections - Schaffer

The Examiner has rejected claims 15-16 under 35 U.S.C. §102(b) as being anticipated by Schaffer (U.S. Patent No. 5,817,913). The Examiner's rejections are respectfully traversed. Claims 15 and 16 have now been amended.

The Examiner points out that the claims are drawn to a Lycopersicon esculentum species characterized by a "capability" of natural dehydration and that plants having such a capability are also taught by U.S. Patent No. 5,817,913.

Schaffer, much in the same as Eshed and Zamir failed to identify the wrinkling trait of the IL4-4 introgression line, simply because Schaffer's plants were not grown post fruit ripening. As described above with respect to the Eshed and Zamir rejection, Claims 15 and 16 have now been amended to reflect actual fruit characteristics and not capabilities thereby distinguishing the claimed subject matter from the prior art.

Although these plants, as well as other plants such as those developed by Eshed and Zamir produce fruit which have the "inherent capability" to dehydrate, the prior art nevertheless does not teach of such plants or of methods of producing such plants, nor does it motivate generation of such plants and therefore the prior art does not anticipate nor render obvious the claimed invention.

In view of the above amendments and remarks it is respectfully submitted that claims 1, 2, 5, 11, 15 and 16 are now in condition for allowance. Prompt notice of allowance is respectfully and earnestly solicited.

Respectfully submitted,

Martin O Moynihan

Martin Moynihan,
Registration No. 40,338

Enc.: Appendix A – Figures 1-5c
Articles by: Frary et al.
Nesbitt et al.
Appendix B - Poster Presentations
Declaration and CV by Dr. Arthur Schaffer
Article by Fulton et al.
Three month extension fee

September 1, 2005
Martin Moynihan
PRTSI, Inc.
P.O. Box 16446
Arlington, VA 22215

Tel: (703) 598-7851 Fax: (703) 415-4864

Theor Appl Genet (2004) 108:485–496
DOI 10.1007/s00122-003-1422-x

ORIGINAL PAPER

A. Frary · T. M. Fulton · D. Zamir · S. D. Tanksley

Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *L. pennellii* cross and identification of possible orthologs in the Solanaceae

Received: 21 May 2003 / Accepted: 4 July 2003 / Published online: 22 January 2004

Abstract In this study, the advanced backcross QTL (AB-QTL) mapping strategy was used to identify loci for yield, processing and fruit quality traits in a population derived from the interspecific cross *Lycopersicon esculentum* E6203 × *Lycopersicon pennellii* accession LA1657. A total of 175 BC_2 plants were genotyped with 150 molecular markers and BC_2F_1 plots were grown and phenotyped for 25 traits in three locations in Israel and California, U.S.A. A total of 84 different QTLs were identified, 45% of which have been possibly identified in other wild-species-derived populations of tomato. Moreover, three fruit-weight/size and shape QTLs (*fsz2b.1*, *fw3.1*/*fsz3.1* and *fs8.1*) appear to have putative orthologs in the related solanaceous species, pepper and eggplant. For the 23 traits for which allelic effects could be deemed as favorable or unfavorable, 26% of the identified loci had *L. pennellii* alleles that enhanced the performance of the elite parent. Alleles that could be targeted for further introgression into cultivated tomato were also identified.

Communicated by G. Wenzel

A. Frary
Department of Biology,
Izmir Institute of Technology,
Urla, 35437 Izmir, Turkey

A. Frary · T. M. Fulton · S. D. Tanksley (✉)
Departments of Plant Breeding and Plant Biology,
Cornell University,
Ithaca, NY 14853, USA
e-mail: sdt4@cornell.edu
Tel.: +1-607-2551673
Fax: +1-607-2556683

T. M. Fulton
Institute for Genomic Diversity,
Cornell University,
Ithaca, NY 14853, USA

D. Zamir
The Faculty of Agriculture, Department of Field Crops,
Hebrew University,
Box 12, 76-100 Rehovot, Israel

Introduction

Twenty years ago the first molecular genetic-linkage map of tomato was published (Tanksley et al. 1992). This map was based on an F_2 population derived from an interspecific cross between cultivated tomato, *Lycopersicon esculentum*, and its wild relative, *Lycopersicon pennellii*. Since this initial report, maps for other and more advanced *L. esculentum* × *L. pennellii* populations (for example, Eshed and Zamir 1995; Haanstra et al. 1999) and for populations from other wild species crosses (for example, Goldman et al. 1995; Tanksley et al. 1996; Fulton et al. 1997a; Bernacchi et al. 1998) have been published. Frequently these interspecific populations have also been used for the identification of quantitative trait loci (QTLs) for important agronomic and horticultural traits. As a result, comprehensive QTL information is now available for populations derived from several wild species of tomato: *Lycopersicon hirsutum* (Bernacchi and Tanksley 1997; Bernacchi et al. 1998), *Lycopersicon peruvianum* (Fulton et al. 1997b), *Lycopersicon parviflorum* (Fulton et al. 2000) and *Lycopersicon pimpinellifolium* (Grandillo and Tanksley 1996; Tanksley et al. 1996; Doganlar et al. 2002a). Most of this information was provided by analysis of advanced backcross (AB) populations. Although two studies examined some growth (de Vicente and Tanksley 1993) and yield-related (Eshed and Zamir 1995) parameters in *L. pennellii*-derived F_2 and introgression-line populations, there has been no report of QTLs identified in a *L. pennellii* AB population.

The AB-QTL mapping strategy integrates the processes of QTL discovery and introgression from wild germplasm into elite material (Tanksley and Nelson 1996). Instead of an F_2 population, this approach uses BC_2 or BC_3 populations derived from an interspecific cross for the identification and mapping of trait loci. Thus, both molecular-marker and phenotypic analyses occur at a more advanced generation when the cultivated parent's alleles are at a much higher frequency. Once favorable alleles for various loci are identified, only a few more crosses are required to develop near-isogenic lines

that can be field-tested and used for variety development. The AB-QTL method was first applied in tomato (Tanksley et al. 1996) and has since been adapted for use in rice (Xiao et al. 1996, 1998; Moncada et al. 2001), wheat (Huang et al. 2003), maize (Ho et al. 2003) and pepper (Rao et al. 2003).

The present paper describes the results from an AB-QTL study of a *L. esculentum* (cultivar E6203) × *L. pennellii* (accession LA1657) BC_2/BC_2F_1population. *L. pennellii* is found in some of the most-arid habitats of all tomato species, and accessions within the species can exhibit extreme genetic variability (Rick and Tanksley 1981). Like most other *L. pennellii* accessions, LA1657 is self-incompatible. Both of the previous *L. pennellii* QTL studies used the self-compatible accession, LA716 (de Vicente and Tanksley 1993; Eshed and Zamir 1995). Both LA1657 and LA716 are from the western regions of Peru; however, their distributions are not identical. LA1657 is usually found in northern regions of the geographic distribution while LA716 is found in southern regions. Moreover, LA1657 prefers higher elevations (about 700 m) than LA716 (20 m) (Rick and Tanksley 1981). Because LA1657 is from a different region of the geographical distribution of the species and is genetically divergent from LA716, it was chosen for this study. The AB-QTL population was grown in three locations in two important tomato-producing regions: Israel and California, U.S.A. Plots were assessed for 25 yield, processing and fruit-appearance traits. Thus, this work extends tomato AB-QTL analyses to a fifth wild species and allows more extensive cross-species comparisons of the control of agronomically important traits in tomato and other solanaceous crops.

Materials and methods

Population development and field evaluations

The population was developed using the processing inbred line *L. esculentum* cultivar E6203 (hereafter referred to as LE) as the recurrent parent and *L. pennellii* accession LA1657 (hereafter referred to as PN) as the donor parent. A total of 320 BC_1 plants were derived from a single F_1 individual and were genotyped with several RFLP markers to select against undesirable phenotypes. TG125 was used to select for homozygous LE alleles at the self-incompatibility locus, *S*, on chromosome 1 to increase the fertility of the plants. TG167 and TG36 were used to screen for LE alleles at fruit-weight QTLs on chromosomes 2 (*fw2.2*) and 11 (*fw11.3*), respectively, to select for larger fruit. In addition, TG279 was used to select for homozygous LE alleles at the *Sp* locus on chromosome 6, thus ensuring that the plants would have a determinate growth habit. This type of growth habit is essential for mechanical harvesting of processing tomatoes. After this marker-assisted selection, eight BC_1 plants were backcrossed to LE to obtain 175 BC_2 plants which were genotyped with RFLP markers for map development. BC_2F_1 families were derived from each of the BC_2 individuals by crossing with TA496 (E6203+*Tm2*^{a}) and were field-tested during the summer of 1998 in Akko, Israel (IS), Woodland, California, U.S.A. (CA1) and Acampo, California, U.S.A (CA2). Plants were grown in randomized plots of 30 plants each with six plots of LE as controls.

Marker and linkage analysis

Genomic DNA extraction, restriction enzyme digestion, Southern hybridization, washing and autoradiography were performed as described in Bernatzky and Tanksley (1986). Parental DNA was surveyed for polymorphism after digestion with *Eco*RI and *Hin*dIII using RFLP markers that were selected at 3-cM intervals from the high-density tomato map (Tanksley et al. 1992). From the surveys, 150 polymorphic markers spanning the entire genome at intervals of less than 20 cM were chosen to genotype the BC_2 individuals.

Marker segregation was tested for significant (P<0.001) deviation from the expected frequency of heterozygotes for a BC_2 population (25%) using the χ^2 goodness-of-fit analysis. The "group" and "ripple" commands of Mapmaker (Lander et al. 1987) were used to establish the most-likely order of markers in each linkage group at LODs 4.0 and 3.0, respectively. Recombination was computed in Kosambi units (Kosambi 1944) using the QGene computer program (Nelson 1997).

Trait evaluations

A total of 25 agronomic traits were evaluated for each plot. Six of the traits were measured at all three locations, seven at two locations and the remaining 12 at only one location. The criteria used for assessing each trait are described below.

Yield traits

Total yield (YLD), red yield (RDY) and percent green yield (PGY) were measured in both IS and CA1. YLD was measured in kilograms and pounds, respectively, and included both ripe (red) and unripe (green) fruit. RDY was the weight of the ripe-red fruit and the weight of the unripe fruit was used to calculate PGY. Plant fertility (FERT) was evaluated only in CA2 using a scale of 1 to 5. A low-fertility rating indicated that the plot had reduced fruit set while a high rating indicated heavy fruit set. The percentage of rotten fruit on the plants in a plot (ROT) at harvest time was assessed only in IS.

Processing traits

Soluble solids content (SSC) was measured in all three locations in Brix using a refractometer as described in Tanksley et al. (1996). Higher values indicated increased sugar content. Soluble solids content was multiplied by red yield to obtain Brix × red yield (BRY) in IS and CA1. This value gives an estimate of the amount of processed product that can be expected from a given line. Juice viscosity (VIS) was measured as Bostwick only in CA1, lower values indicated greater viscosity. Fruit pH (PH) was also measured only in CA1. Thickness of the fruit pericarp (PCP) was evaluated on transverse sections of the fruit on a scale of 1 to 5 (1, thin; 5, thick pericarp) in IS and CA2, and in millimeters in CA1. Fruit firmness (FIR) was assessed by hand-squeezing the fruit (1, soft; 5, very firm). Stem retention (STR) was evaluated only in IS as the percentage of fruit that retained their stems after harvest by shaking the fruit from the plants.

Fruit appearance

Fruit weight (FW) in grams was measured on a random sample of approximately 50 fruit from each plot in IS and CA1. In CA2, fruit size (FSZ) was rated visually (1, very small; 5, very large). Fruit shape (FS) was also measured visually in all three locations on a scale of 1 to 5 where 1 indicated round fruit and 5 indicated elongated fruit.

Fruit color was assessed in four ways. The external color (EC) of ripe fruit was measured using a scale of 1 to 5 (1, light-red; 5, dark-red) in all three locations. Internal color (IC) was also

measured on transverse sections of the fruit in all locations using the same scale as EC. The amount of orange coloration (OR) on the fruit exterior was measured only in CA2 using a scale of 1 to 5:1, very orange; 5, very red. Fruit color (FC) was also measured on raw, de-aerated puree using a spectrophotometer in CA1.

Puffiness (PUF), or the amount of intralocular air space in transversely cut fruit, was evaluated in IS and CA1 using a scale of 1 to 5 (1, very puffy; 5, not puffy). Epidermal reticulation (ER) was measured in IS and CA2, and described whether the fruit skin was smooth (scored as 1) or reticulated like a cantaloupe (scored as 5). The percentage of the fruit that were cracked (PCF) was evaluated only at CA1. Yellow eye (YE) assessed whether the stem-scar penetrated into the fruit. This was measured in CA1 by examining longitudinally cut fruit and estimating the percentage of fruit with YE. Grey wall (GW) was measured in CA1 on transversely cut fruit and was also assessed as a percentage of the fruit with GW. The color of the gel (GG) in the interior of the fruit was scored in CA2 using a scale of 1 (green-gel) to 2 (red-gel).

Data analysis

Pearson's correlation coefficients were calculated for each trait/location combination using the QGene program (Nelson 1997). QGene was also used to perform single-point regression analysis to identify molecular markers with significant linkage to each trait. A QTL is only reported here if it was observed in two or more locations at $P<0.01$ or in one location at $P<0.001$. The percent of the phenotypic variation explained (%PVE) by a given QTL was calculated from the regression of each marker/phenotype combination. The percent phenotypic change or additivity (%A) associated with the presence of a PN allele at a given locus was calculated as $2\times100[(AB-AA)/AA]$, where AA was the phenotypic mean for individuals homozygous for the LE allele at the most-significant marker for the locus and AB was the mean for heterozygous individuals. Because half of the individuals in each BC_2F_1 plot would be heterozygous for any fragment that was heterozygous in the BC_2 generation, the factor of 2 was included to obtain the estimate of %A. Multiple regression analysis was performed in StatView (SAS Institute Inc., Cary, N.C.).

Results

Marker segregation and the genetic map

A total of 152 RFLP markers were genotyped for the BC_2 population. Of these, 110 (72%) were segregating and could be mapped; the remaining 42 markers were fixed for LE alleles. Many of the markers fixed for LE alleles corresponded to the chromosomal regions for which marker-assisted selection was applied to remove the wild parent allele in the BC_1 population. Thus, the top half of chromosome 1 was fixed for LE alleles as a result of selection at the *S* locus (TG125). Marker-assisted selection at *fw2.2* and *fw11.3* resulted in fixation of the middle of chromosome 2 and the bottom half of chromosome 11. In addition, selection at the *Sp* locus on chromosome 6 resulted in fixation of the middle part of this chromosome for LE alleles. Three other regions of the genome encompassing more than one marker were also fixed for LE alleles: the bottom of chromosome 1, a bottom portion of chromosome 4 and the top of chromosome 7. Fixation of these regions cannot be explained by marker-assisted selection. Instead, it may be the result of genetic drift because the BC_1 population that gave rise to the BC_2 was very small.

The average number of heterozygotes per locus was 27% which was nearly identical to the expected value of 25% for a BC_2 population. A total of 31 markers (28%) showed significant ($P<0.001$) deviation from the expected frequency of heterozygotes. Most of these markers were concentrated on six chromosomes. Three regions showed severe skewing with fewer heterozygotes than expected: the bottom of chromosome 5 (TG 60 to CT138, three markers), the top of chromosome 6 (CT216 to TG178, two markers) and the top of chromosome 11 (TG497 to TG523, three markers). Three larger chromosomal regions exhibited segregation distortion with an excess of heterozygotes. The bottom half of chromosome 7 (TG217 to CT195, six markers) and the top half of chromosome 10 (TG230 to TG408, five markers) were moderately skewed while the top half of chromosome 12 (TG180C to CT211A, five markers) was very severely distorted. More than 90% of the population was heterozygous for three of the markers (TG180C, TG68, TG263) in this region.

The 110 mapped markers fell into 15 linkage groups, as markers from the tops and bottoms of chromosomes 2, 4 and 6 could not be linked at the LOD 3.0 threshold (Fig. 1). In all, 87 (79%) of the markers were considered to be framework markers as they were positioned with a ripple at $LOD\geq3.0$. All but one of the remaining markers mapped to the intervals between framework markers at $2.0\leq LOD<3.0$. TG587 on linkage group 4 did not link to the rest of the linkage group, therefore it was assigned to a separate linkage group. The map spanned approximately 703 cM, 55% of the genetic distance encompassed by the high-density tomato map (Tanksley et al. 1992). Coverage was primarily limited by the high percentage of non-segregating markers (28%) many of which (23 of 41 markers) corresponded to regions that were affected by marker-assisted selecton. With only one exception (TG581 on the bottom of chromosome 6), the marker order of the framework map agreed with the high-density map.

Trait correlations

For traits measured in more than one location, the strongest correlations across locations were observed for YLD and RDY ($r=0.72$ and 0.71, respectively) in IS and CA1, and the yield-derived trait, BRY ($r=0.61$), at the same locations (data not shown). FW/FSZ also showed significant ($P<0.05$) positive correlations ($r=0.31$ to 0.53) across locations as did ER ($r=0.47$), SSC ($r=0.19$ to 0.33) and EC ($r=0.23$ to 0.27). None of the other traits that were measured in more than one location (FS, IC, PCP, FIR, PUF and PGY) showed significant correlations across locations.

Within each location, significant correlations were also detected between many traits. However, only those that were highly significant ($P<0.001$), or were observed in more than one location, are described here. For traits

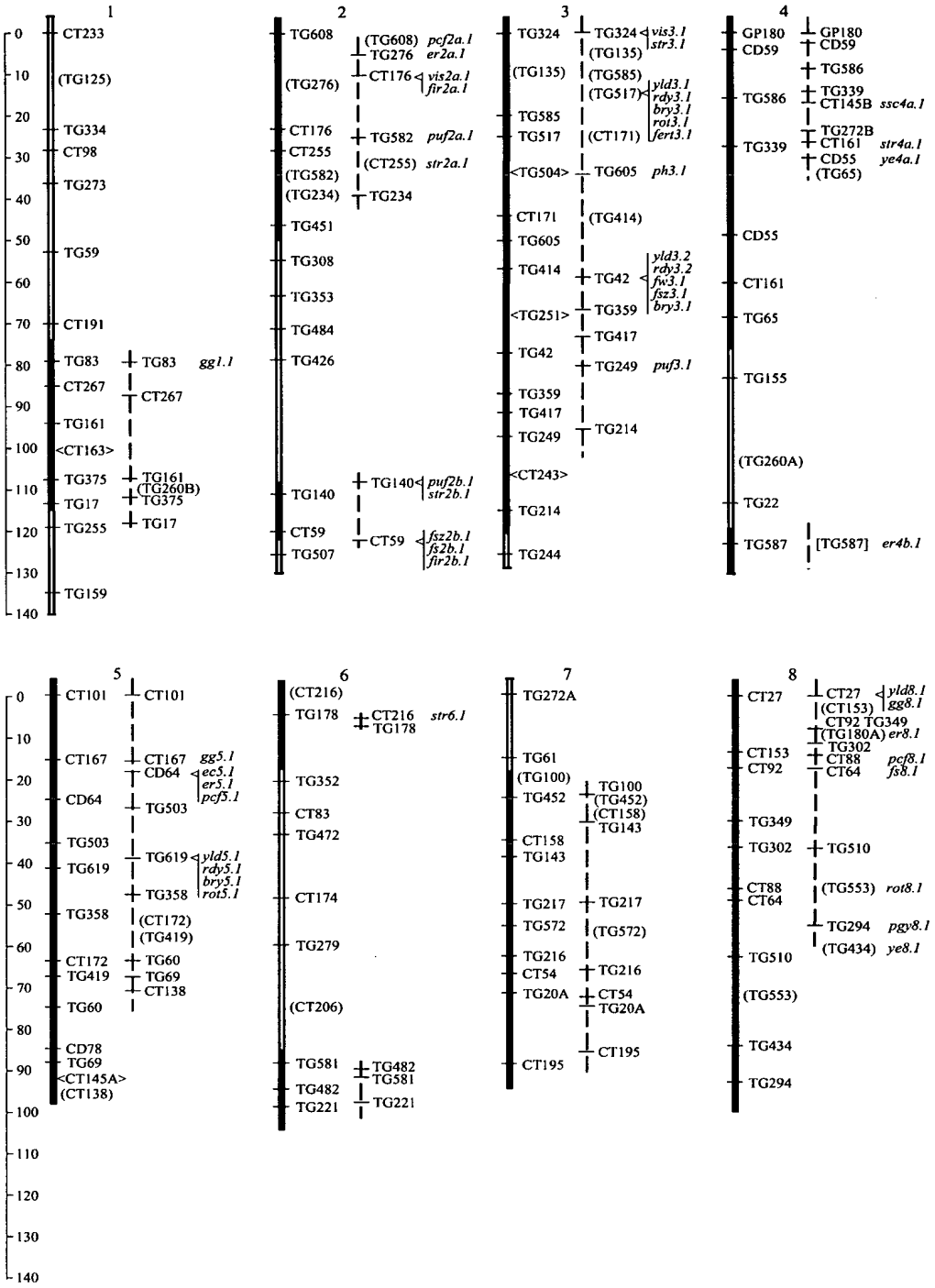
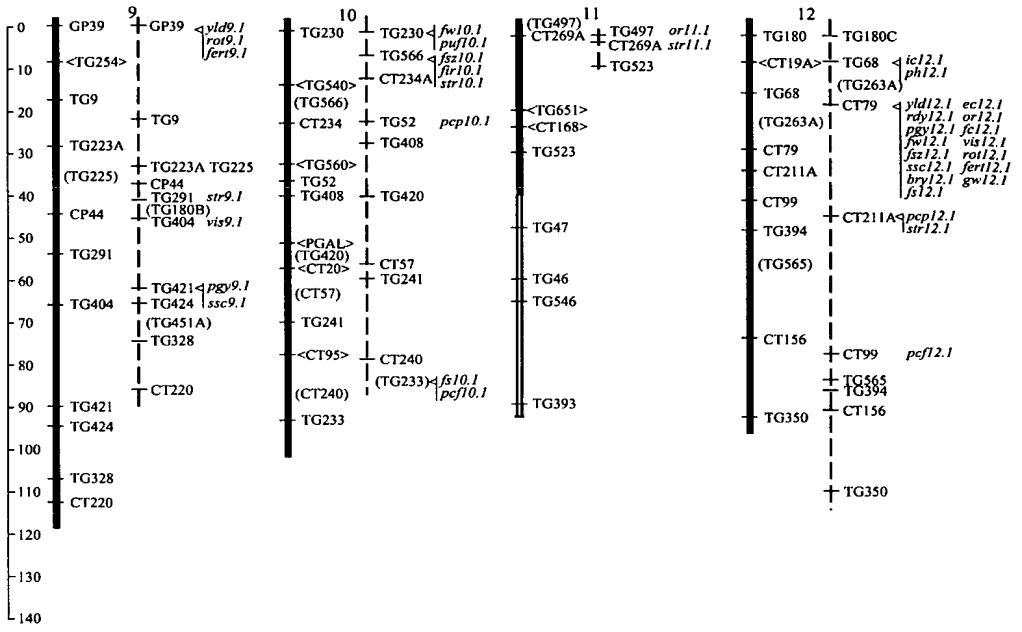


Fig. 1 Comparison of the BC_2 molecular linkage groups (*dashed lines*) to the corresponding chromosomes from the high-density tomato map (*bars*). The centiMorgan scale is given on the far left. Sections of the chromosomes drawn as *open bars* were not segregating for PN alleles in the population. Markers in "<>" were also fixed for LE alleles. Markers in"()" were ordered at LOD<3.0. The marker in "[]" did not link to the rest of the linkage group. Identified QTLs are shown to the right of the BC_2 linkage groups. Co-segregating QTLs are connected by a *vertical line* with a "<" to indicate the most-significant marker

measured in multiple locations, correlation values are averaged across locations. YLD and RDY were both positively correlated with FW (r=0.55 and 0.53, respectively) but were negatively associated with SSC (r=–0.46 and –0.45, respectively). Negative correlations were also observed between SSC and the traits FW/FSZ (r=–0.46), VIS (r=–0.68), EC (r=–0.23) and FERT (r=–0.28). FW/FSZ was positively correlated with EC in two locations (r=0.31), FERT in CA2 (r=0.40) and YE in CA1 (r=0.46). The fruit-color traits, EC and IC, were positively associated in all three locations (r=0.53).

Fig. 1 (continued)



QTLs detected for each trait

A total of 48 QTLs were identified for the 15 traits measured in IS, 43 (90%) of these were significant at $P<0.001$ and the remaining five were significant at $P<0.01$ and identified in at least one other location. A total of 39 QTLs were detected for the 18 traits measured in CA1 with 28 (72%) of the loci detected at $P<0.001$. In addition, 25 QTLs were identified for the 11 traits measured in CA2 with 24 (96%) of the loci identified at the more-stringent significance threshold. For many traits, the QTLs detected in different locations mapped to the same chromosomal positions. When these overlapping QTLs are counted as single loci, it is found that 84 different loci were identified for the 25 traits measured in the study. QTLs were detected on all chromosomes except chromosome 7 with the most loci found on chromosome 12, 20 QTLs. The QTL identified for each trait are described below, listed in Table 1 and mapped in Fig. 1.

Yield traits

Six QTL were identified for total yield on five different chromosomes. All six of these loci were detected in both IS and CA1. The locus on chromosome 12, *yld12.1*, was the most-significant and explained 56% and 32% of the phenotypic variation for the trait in CA1 and IS, respectively. For the other QTL, the percent phenotypic-variance explained (%PVE) was 10% or less. As determined by multiple regression analysis, together the six loci explained 39 and 28% of the variation for yield in CA1 and IS, respectively. The PN allele of only *yld9.1*, which was a relatively minor QTL in terms of significance and magnitude of effect, was associated with increased yield. Four QTLs were detected for red yield on three different chromosomes and all four were identified in both locations where this trait was measured. All of the QTLs had small %PVEs except for *rdy12.1* which controlled as much as 61% of the variance for red yield (in CA1). Overall, the four RDY loci accounted for 34% of the red yield variation in CA1. None of the QTLs showed favorable effects from the wild-alleles.

Three QTLs were found for the percent green yield on chromosomes 8, 9 and 12. Although the loci were highly significant ($P<0.0001$), none of them was detected in both IS and CA1, the two locations where the trait was measured. The most-significant QTL, *pgy12.1*, accounted for up to 19% of variation for the trait. For all three loci, the LE alleles were associated with an increased percent green yield. Three loci were also identified for fertility. As with YLD, RDY and PGY, the locus on chromosome 12, *fert12.1*, had the greatest magnitude of effect, a PVE of 41%. The three FERT loci accounted for 24% of the variation for the trait at CA2. For one of the QTLs, *fert 9.1*, the PN allele was associated with increased fertility. Five QTLs were detected for the amount of rotten fruit on the plants. Similar to the other yield traits, the locus on chromosome 12, *rot12.1*, was the most-significant and explained 19% of the variation in the trait. The other QTLs each accounted for less than 10% of the PVE. Together, the five ROT loci explained 24% of the variation for the amount of rot in IS. For all but one locus, *rot9.1*, the wild-alleles were associated with an increase in the amount of rotten fruit.

Table 1 Putative QTLs identified for each trait. *P*-values for the most-significant marker for each locus are given for Israel (IS) and the two California locations (CA1 and CA2). *Nd* indicates that the trait was not determined at that location, *ns* indicates that the marker was not significant. The percent phenotypic variation explained (%PVE) and percent additivity (%A) are only given for the location for which the QTL was most-significant (indicated by *P*-value in bold). The favorable-allele column indicates whether the *L. esculentum* (LE) or *L. pennellii* (PN) allele was associated with an agronomically favorable effect on the trait. The relative significance of each QTL is coded such that the number of +s indicates the number of locations at which the QTL was detected at $0.001<P<0.01$ and the number of *s indicates the number of locations at which the QTL was identified at $P<0.001$. Populations with putative orthologs are abbreviated: CA = *Capsicum annuum*, pepper; CM = *Lycopersicon cheesmanii*; H = *L. hirsutum*; PF = *L. parviflorum*; PM = *L. pimpinellifolium*; PN = *L. pennellii*; PV = *L. peruvianum*; SM = *Solanum melongena*, eggplant. CA1 = intra-specific *C. annuum* F_2 population (Ben Chaim et al. 2001), CA2 = *C. annuum* x *C. frutescens* advanced backcross population (Rao et al. 2003), CM1 = *L. esculentum* x *L. cheesmanii* F_2 population (Paterson et al. 1991); CM2 = *L. esculentum* x *L. cheesmanii* recombinant inbred population (Goldman et al. 1995); CM3 = *L. esculentum* x *L. cheesmanii* F_2 population (Monforte et al. 1997); H1 = *L. esculentum* x *L. hirsutum* advanced backcross population (Bernacchi et al. 1998); H2 = *L. esculentum* x *L. hirsutum* near-isogenic lines (Monforte et al. 2001); PF = *L. esculentum* x *L. parviflorum* advanced backcross population (Fulton et al. 2000); PM1 = *L. esculentum* x *L. pimpinellifolium* advanced backcross population (Tanksley et al. 1996); PM2 = *L. esculentum* x *L. pimpinellifolium* backcross population (Grandillo and Tanksley 1996); PM3 = *L. esculentum* x *L. pimpinellifolium* F_2 population (Monforte et al. 1997), PM4 = *L. esculentum* x *L. pimpinellifolium* backcross population (Chen et al. 1999); PM5 = *L. esculentum* x *L. pimpinellifolium* F_2 population (Lippman and Tanksley 2001); PM6 = *L. esculentum* x *L. pimpinellifolium* advanced backcross population (Fulton et al. 2002); PM7 = *L. esculentum* x *L. pimpinellifolium* inbredbackcross lines (Doganlar et al. 2002a); PN = *L. esculentum* x *L. pennellii* introgression lines (Eshed and Zamir 1995); PV1 = *L. esculentum* x *L. peruvianum* advanced backcrosspopulation (Fulton et al. 1997b); PV2 = *L. esculentum* x *L. peruvianum* advanced backcrosspopulation (Fulton et al. 2002); PV3 = *L. esculentum* x *L. peruvianum* near-isogenic lines (unpublished data); SM = *Solanum linnaeanum* x *S. melongena* F_2 population (Doganlar et al. 2002b)

Trait	QTL	Chr	Marker	*P*-value			%PVE	%A[a]	Favorable allele[b]	Relative significance	Populations with putative orthologs
				IS	CA1	CA2					
Total yield	*yld3.1*	3	TG517	**0.0003**	0.002	nd	8	−71	LE	*+	
	yld3.2	3	TG42	**0.0003**	0.002	nd	8	−77	LE	*+	H1;PF;PN
	yld5.1	5	TG619	**<0.0001**	0.0003	nd	10	−84	LE	**	
	yld8.1	8	CT27	**0.002**	0.004	nd	6	−78	LE	++	PF;PV1
	yld9.1	9	GP39	**0.005**	0.004	nd	5	86	PN	++	
	yld12.1	12	CT79	<0.0001	**<0.0001**	nd	56	−128	LE	**	H1;PV1
Red yield	*rdy3.1*	3	TG517	**0.0009**	0.005	nd	7	−70	LE	*+	
	rdy3.2	3	TG42	**0.0003**	0.005	nd	8	−80	LE	*+	
	rdy5.1	5	TG619	**0.0001**	0.001	nd	9	−84	LE	*+	PF;PV1
	rdy12.1	12	CT79	<0.0001	**<0.0001**	nd	61	−114	LE	**	H1
Percent green yield	*pgy8.1*	8	TG294	**0.0001**	ns	nd	9	271	LE	*	PV1
	pgy9.1	9	TG421	**<0.0001**	ns	nd	12	313	LE	*	H1
	pgy12.1	12	CT79	ns	**<0.0001**	nd	19	509	LE	*	
Fertility	*fert3.1*	3	TG517	nd	nd	**0.0002**	9	−53	LE	*	
	fert9.1	9	GP39	nd	nd	**0.0007**	7	72	PN	*	
	fert12.1	12	CT79	nd	nd	**<0.0001**	41	−97	LE	*	
Rotten	*rot3.1*	3	TG517	**0.0003**	nd	nd	8	−81	PN	*	
	rot5.1	5	TG619	**0.0003**	nd	nd	7	−81	PN	*	
	rot8.1	8	TG553	**0.0002**	nd	nd	9	−80	PN	*	PM7
	rot9.1	9	GP39	**0.0008**	nd	nd	7	121	LE	*	
	rot12.1	12	CT79	**<0.0001**	nd	nd	19	−106	PN	*	
Soluble solids content	*ssc4a.1*	4a	CT145b	ns	**0.0006**	ns	8	−15	LE	*	PF;PM1,6
	ssc9.1	9	TG421	ns	**0.0006**	ns	8	15	PN	*	H1;PF;PM7;PN
	ssc12.1	12	CT79	**<0.0001**	<0.0001	0.0005	30	48	PN	***	PM1,4;PN;PV2
Brix × red yield	*bry3.1*	3	TG517	**0.0003**	ns	nd	8	−65	LE	*	
	bry3.2	3	TG42	**0.0003**	0.006	nd	8	−68	LE	*+	H1;PN
	bry5.1	5	TG619	**<0.0001**	0.002	nd	11	−79	LE	*+	PF;PV1
	bry12.1	12	CT79	<0.0001	**<0.0001**	nd	53	−126	LE	**	H1;PN;PV1
Viscosity	*vis2a.1*	2a	CT176	nd	**0.0009**	nd	7	19	LE	*	
	vis3.1	3	TG324	nd	**0.0004**	nd	8	−20	PN	*	
	vis9.1	9	TG404	nd	**<0.0001**	nd	18	−24	PN	*	PM1;PV1
	vis12.1	12	CT79	nd	**<0.0001**	nd	18	−30	PN	*	
pH	*ph3.1*	3	TG605	nd	**0.0007**	nd	8	4	PN	*	
	ph12.1	12	TG68	nd	**<0.0001**	nd	18	7	PN	*	PM4;PV1
Pericarp thickness	*pcp10.1*	10	TG52	ns	ns	**<0.0001**	17	−24	LE	*	PF;PM2
	pcp12.1	12	CT211	**0.001**	ns	0.007	6	16	PN	++	
Fruit firmness	*fir2a.1*	2a	CT176	**<0.0001**	ns	ns	16	−50	LE	*	
	fir2b.1	2b	CT59	ns	ns	**<0.0001**	13	−44	LE	*	PM1
	fir10.1	10	TG566	ns	ns	**0.0006**	9	−31	LE	*	PF;PM7

Table 1 (continued)

Trait	QTL	Chr	Marker	*P*-value			%PVE	%A[a]	Favorable allele[b]	Relative significance	Populations with putative orthologs
				IS	CA1	CA2					
Stem retention	*str2a.1*	2a	CT255	**0.0001**	nd	nd	10	55	LE	*	
	str2b.1	2b	TG140	**<0.0001**	nd	nd	17	102	LE	*	H1;PV1
	str3.1	3	TG324	**<0.0001**	nd	nd	11	65	LE	*	
	str4a.1	4a	CT161	**<0.0001**	nd	nd	15	73	LE	*	
	str6a.1	6a	CT216	**0.0001**	nd	nd	9	69	LE	*	
	str9.1	9	TG291	**0.0002**	nd	nd	8	44	LE	*	H1;PV1
	str10.1	10	TG566	**0.0003**	nd	nd	8	42	LE	*	H1;PF;PM2
	str11.1	11	CT269	**<0.0001**	nd	nd	25	139	LE	*	
	str12.1	12	CT211	**0.0002**	nd	nd	8	−35	PN	*	
Fruit-weight	*fw3.1*	3	TG42	**0.0002**	0.004	nd	8	−26	LE	*+	CM1,2;PF;PM1,4,5;PN;CA1,2;SM
	fw10.1	10	TG230	**0.0002**	0.004	nd	8	−20	LE	*+	PF;PN;PV1
	fw12.1	12	CT79	**<0.0001**	<0.0001	nd	20	−30	LE	**	PM4;PV1
Fruit size	*fsz2b.1*	2b	CT59	nd	nd	**0.0006**	8	−24	LE	*	CM1,3;H1;PF;PM1,3,5;PN;PV;CA1,2
	fsz3.1	3	TG42	nd	nd	**0.0008**	8	−23	LE	*	CM1,2;PF;PM1,4,5;PN;CA1,2
	fsz10.1	10	TG566	nd	nd	**<0.0001**	18	−29	LE	*	PF;PN;PV1
	fsz12.1	12	CT79	nd	nd	**<0.0001**	16	−32	LE	*	PM4;PV1
Fruit shape	*fs2b.1*	2b	CT59	ns	0.001	**0.0006**	8	−16	LE	*+	PF;PM4;PV1
	fs8.1	8	CT64	**<0.0001**	0.0001	ns	12	−26	LE	**	H1;PF;PM2;PV1;CA1
	fs10.1	10	TG233	0.007	ns	**<0.0001**	17	−21	LE	*+	
	fs12.1	12	CT79	ns	ns	**0.0002**	10	−19	LE	*	PM4;PV1
Fruit external color	*ec5.1*	5	CD64	ns	ns	**0.0001**	12	−48	LE	*	H1
	ec12.1	12	CT79	<0.0001	0.0001	**<0.0001**	16	−44	LE	***	PF;PV1
Fruit internal color	*ic12.1*	12	TG68	**0.0002**	ns	ns	8	−48	LE	*	PF
Fruit orange color	*or11.1*	11	TG497	nd	nd	**0.0004**	10	86	LE	*	
	or12.1	12	CT79	nd	nd	**<0.0001**	20	142	LE	*	
Fruit color (lab)	*fc12.1*	12	CT79	nd	**<0.0001**	nd	18	−45	LE	*	PV1
Puffiness	*puf2a.1*	2a	TG582	0.009	**0.003**	nd	6	90	PN	++	
	puf2b.1	2b	TG140	**<0.0001**	ns	nd	10	47	PN	*	
	puf3.1	3	TG249	**0.0009**	ns	nd	7	25	PN	*	
	puf10.1	10	TG230	**0.0002**	ns	nd	8	24	PN	*	
Epidermal reticulation	*er2a.1*	2a	TG276	**0.0001**	nd	ns	9	87	LE	*	
	er4b.1	4b	TG587	**<0.0001**	nd	<0.0001	43	222	LE	**	H2;PF;PV3
	er5.1	5	CD64	0.002	nd	**0.0007**	9	47	LE	*+	
	er8.1	8	TG180a	**<0.0001**	nd	<0.0001	12	108	LE	**	
Percent cracked fruit	*pcf2a.1*	2a	TG608	nd	**<0.0001**	nd	10	133	LE	*	
	pcf5.1	5	CD64	nd	**0.0003**	nd	8	137	LE	*	
	pcf8.1	8	CT88	nd	**0.0009**	nd	7	118	LE	*	
	pcf10.1	10	TG233	nd	**<0.0001**	nd	13	144	LE	*	
	pcf12.1	12	CT99	nd	**0.0001**	nd	10	−84	PN	*	
Yellow eye	*ye4a.1*	4a	CD55	nd	**0.0006**	nd	11	87	LE	*	
	ye8.1	8	TG434	nd	**0.0001**	nd	13	−80	PN	*	
Grey wall	*gw12.1*	12	CT79	nd	**0.0004**	nd	10	−77	PN	*	
Green gel	*gg1.1*	1	TG83	nd	nd	**0.0003**	10	64	LE	*	
	gg5.1	5	CT167	nd	nd	**0.0001**	11	66	LE	*	
	gg8.1	8	CT27	nd	nd	**0.0001**	12	64	LE	*	PF

[a] %A=200(AB−AA)/AA where AA is the phenotypic mean for individuals homozygous for the *L. esculentum* allele at the most-significant marker and AB is the mean for heterozygous individuals

[b] For pH and fruit shape, this column indicates which allele was associated with an increase in the trait mean

Processing traits

The soluble-solids content of the fruit was determined in all three locations and three different loci were identified. Two relatively minor QTLs mapped to chromosomes 4 and 9, and the most-significant locus, *ssc12.1*, mapped to chromosome 12. This QTL accounted for up to 30% of the variation in soluble solids (in IS). Overall, the three loci accounted for 24% of the variation in the trait in CA1. The PN alleles for both *ssc9.1* and *ssc12.1* were associated with increased soluble solids. Four QTLs were detected for the derived-trait soluble solids (Brix) × red yield and were distibuted on three chromosomes: 3, 5 and 12. By far, the most-significant was *bry12.1* on chromosome 12 which explained 19% of the phenotypic variation in CA1 and 53% in IS. Together, the four BRY loci explained 23% of the Brix × red yield variation in IS. For all four loci, the LE alleles were associated with increased BRY.

The viscosity of juice from the tomatoes was measured only in CA1 where four QTLs were identified. Two of these loci, *vis9.1* and *vis12.1*, each explained 18% of the variation in juice viscosity and all together, the four QTLs accounted for 21% of the phenotypic variation. For all but *vis2a.1*, the wild-alleles had favorable effects and were associated with a more-viscous product. The pH of the fruit was also only determined in CA1. Two loci were detected, *ph3.1* and *ph12.1*, which accounted for 8 and 18% of the PVE for the trait, respectively. For both QTLs, the PN alleles were associated with increased acidity of the fruit.

Pericarp thickness was measured in all three locations and two different QTLs were identified. The more-significant locus was located on chromosome 10 and accounted for 17% of the variation for the trait. The combined effects of both loci explained 14% of the variation for pericarp thickness. The wild alleles for the two loci had opposite effects. The PN allele increased pericarp thickness for *pcp12.1* and decreased it for *pcp10.1*. Fruit firmness was also determined in all three locations and three loci were identified: *fir2a.1*, *fir2b.*1 and *fir10.1*. *Fir2a.1* had the greatest %PVE, 16%. Together, the two loci identified in CA2 accounted for 15% of the variation in firmness. The LE alleles for all three QTLs were associated with firmer fruit. Stem retention was measured only in IS where nine QTLs were detected, the most for any trait in this study. These loci were distributed on eight different chromosomes with two QTLs on the separate linkage groups representing chromosome 2. Most of the loci had magnitudes of effect of 8 to 15%; however, the most-significant QTL, *str11.1*, explained 25% of the variation for stem retention. Overall, the nine loci explained 25% of the variation in the trait. With only one exception, *str12.1*, the LE alleles were associated with decreased stem retention.

Fruit appearance traits

Fruit size was assessed by weighing the fruit in IS and CA1 (FW), and with a visual scale in CA2 (FSZ). Three loci were detected for FW on chromosomes 3, 10 and 12. The QTL on chromosome 12, *fw12.1*, was the most-significant and explained as much as 20% of the PVE. Together, the three FW loci accounted for 12% of the variation for the trait in IS. Four QTLs were identified for FSZ, three of which corresponded closely to the FW loci. The fourth locus was identified on the lower portion of chromosome 2. The FSZ loci on chromosomes 10 and 12 were both highly significant and had similar magnitudes of effect, 18 and 16%, respectively. The combined effects of these four loci explained 15% of the phenotypic variation. It should be noted that marker-assisted selection was deliberately applied to remove three regions containing some of the most-significant fruit-weight QTLs previously identified in tomato: *fw1.1* near the *S* locus on chromosome 1, *fw2.2* on chromosome 2 and *fw11.3* on chromosome 11. Thus, the analysis for fruit-weight loci probably does not reflect the entire potential of this accession of *L. pennellii* as a source of the fruit-weight QTL. For all of the FW and FSZ loci, the PN alleles were associated with reduced fruit size as expected.

Fruit shape was controlled by four QTLs all of which were detected in two of the three locations where the trait was measured. *Fs8.1* had the highest significance levels; however, *fs10.1* had a larger effect on variation for fruit shape, maximums of 12 and 17%, respectively. The three loci detected in CA2 had a combined magnitude of effect of 18%. As expected based on the parental phenotypes, the PN alleles were associated with rounder fruit.

Fruit color was measured in four ways: external color (EC), internal color (IC), the amount of external orange color (OR) and a laboratory measurement on juice (FC). Two QTLs were identified for EC, *ec5.1* and *ec12.1*, accounting for 12 and 16% of the variation for the trait, respectively. Loci for IC were not identified in either CA location; however, one QTL was detected in IS, *ic12.1*. This locus only explained 8% of the phenotypic variation in internal fruit color. Two loci for OR were found in CA2. The more significant QTL, *or12.1*, had a magnitude of effect of 20%. Together, the two loci accounted for 16% of the variation for OR. Only one QTL was identified for FC, *fc12.1*, which explained 18% of the variation in the trait. For all of the fruit-color loci, the LE alleles were associated with improved, that is, redder color.

Puffiness or the amount of air space in the fruit locules was measured in two locations (IS and CA1) where four different QTLs were identified. Two of these QTLs mapped to the different linkage groups of chromosome 2 and the other two loci were located on chromsomes 3 and 10. All of the loci had relatively minor %PVEs of 10% or less and a combined magnitude of effect of 15%. The PN alleles were always associated with decreased puffiness. Epidermal reticulation describes the cantaloupe-like veining that is observed on the skin of some fruit. Four QTLs

controlling this trait were identified in IS and CA2 on chromosomes 2, 4, 5 and 8. The locus on chromsome 4, *er4b.1*, was the most-significant and accounted for as much as 43% of the phenotypic variation in IS. In combination, the four loci had a PVE of 41%. For all four QTLs, the PN alleles were linked to increased reticulation. The percent of cracked fruit was only measured in CA1 where five QTLs were found. The most-significant of these was *pcf10.1* with a %PVE of 13% and, together, the loci accounted for 22% of the phenotypic variation for the trait. For only one QTL, *pcf12.1*, the wild alleles were associated with reduced cracking.

Yellow eye measured the penetration of the stem scar in the fruit. Two QTLs were identified for this trait, *ye4a.1* and *ye8.1*, controlling 11 and 13% of the phenotypic variation, respectively. The PN alleles for these loci had opposite effects increasing the percentage of fruit with yellow eye at *ye4.1* and decreasing it at *ye8.1*. Grey wall was measured only in CA1 where only one QTL was detected. This QTL, *gw12.1*, explained only 10% of the variation for the trait and its PN alleles were associated with improved fruit appearance. The color of the gel in cut fruit was assessed only in CA2. At this location, three QTLs were identified on chromosomes 1, 5 and 8, all of which had similar significances and magnitudes of effect ranging between 10 and 12%. None of the loci showed favorable effects from the wild parent-allele.

Discussion

Segregation distortion

A common feature of many interspecific plant populations is distorted segregation. This has been attributed to structural differences or loci that affect gamete transmission in the affected chromosomal regions (Zamir and Tadmor 1986). Six segments of the genome showed significant skewing of marker segregation ratios in the *L. esculentum* × *L. pennellii* BC_2 population. Regions on chromosomes 5, 6 and 11 had excesses of LE alleles while portions of chromosomes 7, 10 and 12 had higher than expected frequencies of PN alleles. The segregation distortion toward the LE genotype seen on chromosomes 6 and 11 was probably the result of the marker-assisted selection in these regions. Deviant segregation for some of these chromosomal regions has been reported in other tomato populations. For example, segregation distortion toward PN alleles of the top of chromosome 10 was observed in two *L. esculentum* × *L. pennellii* F_2 populations (deVicente and Tanksley 1993; Haanstra et al. 1999). Skewing was also detected in *L. hirsutum*, *L. peruvianum* and *L. parviflorum* interspecific populations for an overlapping region; however, in these populations excesses of LE alleles were observed (Bernacchi and Tanksley 1997; Fulton et al. 1997a, 2000). Similar to the current study, deviation from expected segregation ratios with an excess of LE alleles on chromosome 11 was reported in the *L. hirsutum*, *L. peruvianum* and *L. parviflorum* populations (Bernacchi and Tanksley 1997; Fulton et al. 1997a, 2000). It is interesting to note that none of these studies performed marker-assisted selection for this region. In contrast, deVicente and Tanksley (1993) observed that all of the markers on chromosome 11 were skewed toward the PN alleles in the *L. pennellii* F_2 population. The most-dramatic distortion observed in the BC_2 population occurred on a 45 cM portion of chromosome 12. Approximately 90% of the individuals in the population were heterozygous for this region. Zamir and Tadmor (1986) also saw a very marked preference for PN alleles in this region in an F_2 population.

The reasons for such dramatic segregation distortion are largely unknown. The BC_2 population was derived from a very small BC_1 population and therefore was very susceptible to genetic drift. Such drift might account for both fixation and segregation distortion in the population. Pelham (1968) attributed skewing on chromosome 9 of *L. peruvianum* to a gamete promoter gene. Preferential inheritance of the *L. peruvianum* allele in this region was also observed by Fulton et al. (1997a); however, conclusive evidence of a gamete promoter gene on the chromosome has not been reported. Analysis of the mechanism(s) responsible for distorted segregation is difficult as skewed regions vary greatly among species and even among populations derived from the same parent. Preferable inheritance of certain alleles in a given region has practical ramifications as it may necessitate additional backcross generations to achieve a desired level of homozygosity in breeding programs.

Correlations across locations and between traits

Correlations across locations were not significant for six of the 13 traits, measured in more than one location. However, there were strong associations across locations for the yield and yield-derived traits, and moderate correlations for fruit-weight, soluble solids and external color. From an agronomic perspective, these are the most-important traits for processing tomato cultivars. Similar to many previous studies, YLD/RYD and FW were found to be positively correlated (Stevens and Rudich 1978; Stevens 1986; Tanksley et al. 1996; Fulton et al. 1997b; Bernacchi et al. 1998; Fulton et al. 2000). However, both the yield and fruit-weight/size traits were negatively correlated with SSC. This is a well-documented phenomenon that suggests that attempted improvement in soluble solids will be at the expense of yield (Ibarbia and Lambeth 1971; Stevens 1986; Paterson et al. 1991; Tanksley et al. 1996; Fulton et al. 1997b; Bernacchi et al. 1998; Chen et al. 1999; Fulton et al. 2000; Doganlar et al. 2002a). A negative correlation was also identified between SSC and VIS, a result that was expected based on previous work (Stevens 1986, Fulton et al. 2000) and the fact that juice with higher soluble solids is, by its nature, more viscous. The significant positive correlation between FSZ and FERT suggests that an increased

number of fruit is not necessarily associated with a reduction in fruit size. External and internal fruit colors were also positively correlated as has been observed in *L. peruvianum*, *L. parviflorum* and *L. pimpinellifolium* mapping populations (Fulton et al. 1997b, 2000; Doganlar et al. 2002a).

Conservation of loci across environments

Of the 25 traits evaluated in this work, six were measured in all three locations, seven were assessed in two locations and 12 were determined at only one location. For the 13 traits that were evaluated in more than one environment, 43 QTLs were detected. Of these, two loci (5%) were identified at all three locations and 24 loci (56%) were identified at two locations. The only two QTLs identified in all three locations were *ssc12.1* and *ec12.1*. Notably, all of the YLD and RDY loci (six and four QTLs, respectively) were detected in both locations where these traits were measured. Moreover, all three FW loci and three of the four FS, BRY and ER QTLs were identified at two locations. This conservation across locations suggests that locus by environment interactions for these traits are relatively low. Strong conservation of QTLs across locations has been reported previously for several different interspecific AB-QTL tomato populations (Tanksley et al. 1996; Fulton et al. 1997b; Bernacchi et al. 1998; Fulton et al. 2000).

Co-localization of QTLs

The largest cluster of QTLs was on chromosome 12 where CT79 was a significant marker for 15 different loci. Smaller clusters of loci (three or more QTLs) were also present on chromosomes 2, 3, 5, 9 and 10. As expected, similar or related traits tended to be co-localized in the genome. For example, the four RDY QTLs always mapped with the YLD QTL. In addition, YLD loci mapped to the same regions as FERT QTLs on chromosomes 3, 9 and 12, and FW/FSZ QTLs on chromosomes 3 and 12. FW/FSZ loci also were co-localized with FS QTLs on chromosomes 2 and 12. Many of these clusters of related traits may reflect the pleiotropic effects of single loci. However, linkage of genes cannot be ruled out as a possible cause unless additional mapping is performed. For example, many studies have localized both fruit-weight and shape QTLs to the bottom half of chromosome 2 (reviewed in Grandillo et al. 1999). However, recent isolation of *fw2.2* (Frary et al. 2000) and *ovate* (Liu et al. 2002) have demonstrated that there are indeed distinct fruit-weight and shape-loci in this region of the genome.

QTLs with potential for breeding improved tomatoes

Many previous studies in tomato have demonstrated that phenotypically inferior wild species can be a source of agronomically favorable alleles (deVicente and Tanksley 1993; Eshed and Zamir 1995; Grandillo and Tanksley 1996; Tanksley et al. 1996; Fulton et al. 1997b; Bernacchi et al. 1998; Chen et al. 1999; Fulton et al. 2000; Doganlar et al. 2002a). In the present work, 11 (48%) of 23 traits had at least one QTL for which the *L. pennellii* allele had a positive agronomic effect. Traits for which effects were neither favorable nor unfavorable were excluded from this analysis. For example, pH was not included because increases or decreases in this character are not necessarily positive or negative but must be kept within an acceptable range for processing. Overall, 26% of the identified loci (20/78) had wild-alleles that enhanced the agronomic performance of the advanced backcross lines. Even higher percentages of traits with favorable wild-alleles were obtained with *L. peruvianum* (more than 50%, Fulton et al. 1997b), *L. hirsutum* (60%, Bernacchi et al. 1998) and *L. parviflorum* (70%, Fulton et al. 2000).

Some of the loci identified in this study may be targeted for breeding purposes. The *L. pennellii* allele(s) for the overlapping soluble solids and viscosity QTLs on chromosome 9 improved these two traits by 8 and 18%, respectively. Because of the related nature of these traits, it is probable that these effects are due to pleiotropy. The wild-alleles for several loci, centered around CT79 on chromosome 12, also had beneficial effects. The *L. pennellii* allele(s) at this location was (were) associated with a 48% increase in soluble solids content, an 18% improvement in viscosity, 19% and 10% reductions in fruit rot and cracking, respectively, a 16% increase in pericarp thickness and slight decreases in stem retention and grey wall. Unfortunately, cultivated alleles from the same region were also significantly linked to great improvements in total and red yields (56% and 61%, respectively), fertility (41%) and fruit-weight (20%), and lesser increases in external and internal fruit color.

Although it is possible that the multiple effects of this region of chromosome 12 are the result of pleiotropy, the diversity of phenotypes for the QTLs suggests that more than one locus does indeed exist in the neighborhood of CT79. Given the breeding potential of this region, it may be worthwhile to break the linkage between the sugar- and yield-traits so that the *L. pennellii* allele for improved soluble solids can be introgressed into cultivated tomato. This will require additional mapping to verify that the loci are indeed distinct and the screening of large populations for individuals that contain recombinations that break the linkages between the various traits. Such an approach has been used to break linkages between poor yield, low fruit-weight and high soluble solids in a *L. hirsutum* introgression (Monforte and Tanksley 2000), and between orange fruit color and high sugars in a *Lycopersicon chmielewskii* introgression (Frary et al. 2003a).

Loci shared among populations and species

With the addition of the present study, comprehensive QTL analyses are now available for AB populations derived from crosses with five different wild-species of tomato. In addition, the first QTL studies for pepper (Ben Chaim et al. 2001; Rao et al. 2003) and eggplant (Doganlar et al. 2002b; Frary et al. 2003b) have recently been published. This availability allows the identification of loci that are putatively conserved across tomato and its related wild- and crop-species. Of the 84 QTLs identified in this study, 38 (45%) are possibly the same as loci detected in other populations and species (Table 1). QTLs were considered to be potentially orthologous if they mapped to the same 20-cM region of the high-density tomato map (Tanksley et al. 1992). The majority (76%) of the putatively conserved loci were identified in three or more populations derived from different tomato species. In general, the yield-related, soluble solids and fruit size, shape and color traits had the highest proportions of QTLs that had been previously identified. This is probably because these traits have been examined in many studies whereas traits such as the amount of rotten and cracked fruit, puffiness and yellow eye have been examined in very few or no previous studies. The most-frequently identified loci were: *fsz2b.1*, detected in nine tomato populations representing six different species; *fw3.1/ fsz3.1*, identified in seven tomato populations encompassing four different species and *fs8.1*, detected in four tomato populations representing four different species. In addition, three loci appear to have orthologous counterparts outside of tomato. The fruit-weight/size QTL on chromosome 3 and the fruit-shape locus on chromosome 8 have been identified in pepper, *Capsicum annuum* (Ben Chaim et al. 2001; Rao et al. 2003). Moreover, *fsz2b.1* has been identified in both pepper and eggplant, *Solanum melongena* (Ben Chaim et al. 2001; Doganlar et al. 2002b; Rao et al. 2003). Such putative conservation of loci within the genus *Lycopersicon* and across other solanaceous species re-inforces the validity of the shared QTLs and supports the hypothesis that evolution and domestication in the Solanaceae has proceeded via mutations in loci that have been functionally conserved since divergence from a common ancestor (Doganlar et al. 2002b; Frary et al. 2003b).

Acknowledgements We thank Charles Fleck and T. Casey Garvey at Hunts, and Steve Schroeder at Sunseeds, for assistance with the field tests in California. This work was supported in part by grants from the National Research Initiative Cooperative Grants Program (No. 96-35300-3646) and the Binational Agricultural Research and Development Fund (No. US 2427-94).

References

Ben Chaim A, Paran I, Grube RC, Jahn M, van Wijk R, Peleman J (2001) QTL mapping of fruit-related traits in pepper (*Capsicum annuum*). Theor Appl Genet 102:1016–1028

Bernacchi D, Tanksley SD (1997) An interspecific backcross of *Lycopersicon esculentum* × *L. hirsutum*: linkage analysis and a QTL study of sexual compatibility factors and floral traits. Genetics 147:861–877

Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley SD (1998) Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *L. hirsutum*. Theor Appl Genet 97:381–397

Bernatzky R, Tanksley SD (1986) Majority of random cDNA clones correspond to single loci in the tomato genome. Mol Gen Genet 203:8–14

Chen FQ, Foolad MR, Hyman J, St. Clair DA, Beelman RB (1999) Mapping of QTLs for lycopene and other fruit traits in a *Lycopersicon esculentum* × *L. pimpinellifolium* cross and comparison of QTLs across tomato species. Mol Breed 5:283–299

de Vicente MC, Tanksley SD (1993) QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:585–596

Doganlar S, Frary A, Ku H-K, Tanksley SD (2002a) Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589). Genome 45:1189–1202

Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD (2002b) Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. Genetics 161:1713–1726

Eshed Y, Zamir D (1995) An introgression-line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTLs. Genetics 141:1147–1162

Frary A, Nesbitt TC, Frary A, Grandillo S, Van der Knaap E, Cong B, Lui J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) *fw2.2*: a quantitative trait-locus key to the evolution of tomato fruit size. Science 289:85–88

Frary A, Doganlar S, Frampton A, Fulton T, Uhlig J, Yates H, Tanksley SD (2003a) Fine mapping of quantitative trait loci for improved fruit characterisitcs from *Lycopersicon chmielewskii* chromosome 1. Genome 46:235–243

Frary A, Doganlar S, Daunay MC, Tanksley SD (2003b) QTL analysis of morphological traits in eggplant and implications for conservation of gene function during evolution of solanaceous species. Theor Appl Genet 107:359–370

Fulton TM, Nelson JC, Tanksley SD (1997a) Introgression and DNA marker analysis of *Lycopersicon peruvianum*, a wild relative of the cultivated tomato, into *L. esculentum*, followed through three successive backcross generations. Theor Appl Genet 95:895–902

Fulton TM, Beck-Bunn T, Emmatty D, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley SD (1997b) QTL analysis of an advanced backcross of *Lycopersicon peruvianum* to the cultivated tomato and comparisons with QTLs found in other wild species. Theor Appl Genet 95:881–894

Fulton TM, Grandillo S, Beck-Bunn T, Fridman E, Frampton A, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley SD (2000) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *L. parviflorum* cross. Theor Appl Genet 100:1025–1042

Fulton TM, Bucheli P, Voirol E, Lopez J, Petiard V, Tanksley SD (2002) Quantitative trait loci (QTLs) affecting sugars, organic acids and other biochemical properties possibly contributing to flavor, identified in four backcross populations of tomato. Euphytica 127:163–177

Goldman IL, Paran I, Zamir D (1995) Quantitative trait locus analysis of a recombinant inbred line population derived from a *Lycopersicon esculentum* × *L. cheesmanii* cross. Theor Appl Genet 90:925–932

Grandillo S, Tanksley SD (1996) Analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. Theor Appl Genet 92:935–951

Grandillo S, Ku H-M, Tanksley SD (1999) Identifying the loci responsible for natural variation in fruit size and shape in tomato. Theor Appl Genet 99:978–987

Haanstra JP, Wye C, Verbakel H, Meijer-Dekens F, van den Berg P, Odinot P, van Heusden AW, Tanksley S, Lindhout P, Peleman J (1999) An integrated high-density RFLP-AFLP map of tomato based on two *Lycopersicon esculentum* × *L. pennellii* F_2 populations. Theor Appl Genet 99:254–271

Ho J, McCouch SM, Smith M (2003) Improvement of hybrid yield by an advanced backcross QTL analysis in elite maize. Theor Appl Genet 105:440–448

Huang XQ, Coster H, Ganal MW, Roder MS (2003) Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum*). Theor Appl Genet 106:1379–1389

Ibarbia EA, Lambeth VN (1971) Tomato fruit size and quality interrelationships. J Am Soc Hort Sci 96:199–201

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lippman Z, Tanksley SD (2001) Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small fruited wild species *Lycopersicon pimpinellifolium* and *L. esculentum* var. Giant Heirloom. Genetics 158:413–422

Liu J, van Eck J, Cong B, Tanksley SD (2002) A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. Proc Natl Acad Sci USA 99:13302–13306

Moncada P, Martinez CP, Borrero J, Chatel M, Guach H, Guimaraes E, Tohme J, McCouch SR (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa* × *O. rufipogon* BC_2F_2 population evaluated in an upland environment. Theor Appl Genet 102:41–52

Monforte AJ, Tanksley SD (2000) Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome-1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. Theor Appl Genet 100:471–479

Monforte AJ, Asins MJ, Carbonell EA (1997) Salt tolerance in *Lycopersicon* species. V. Does genetic variability at quantitative trait loci affect their analysis? Theor Appl Genet 95:284–293

Monforte AJ, Friedman E, Zamir D, Tanksley SD (2001) Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: deductions about natural variation and implications for germplasm utilization. Theor Appl Genet 102:572–590

Nelson JC (1997) QGENE: software for marker-based genomic analysis and breeding. Mol Breed 3:229–235

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinovitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. Genetics 127:181–197

Pelham J (1968) Disturbed segregation of genes on chromosome 9: gamete promoter, *Gp*, a new gene. Tomato Genet Coop 18:27–29

Rao GU, Ben Chaim A, Borovsky Y, Paran I (2003) Mapping of yield-related QTLs in pepper in an interspecific cross of *Capsicum annuum* and *C. frutescens*. Theor Appl Genet 106:1457–1466

Rick CM, Tanksley SD (1981) Genetic variation in *Solanum pennellii*: comparisons with two other sympatric tomato species. Plant Syst Evol 139:11–45

Stevens MA (1986) Inheritance of tomato fruit quality components. Plant Breed Rev 4:273–311

Stevens MA, Rudich J (1978) Genetic potential for overcoming physiological limitations on adaptability, yield and quality in the tomato. Hort Science 13:673–678

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203

Tanksley SD, Ganal MW, Prince JP, deVicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Bunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor Appl Genet 92:213–224

Xiao J, Li J, Yuan L, Tanksley SD (1996) Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics 140:745–754

Xiao J, Li J, Grandillo S, Ahn S, Yuan L, Tanksley SD, McCouch SR (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899–909

Zamir D, Tadmor Y (1986) Unequal segregation of nuclear genes in plants. Bot Gaz 147:355–358

Comparative Sequencing in the Genus Lycopersicon: Implications for the Evolution of Fruit Size in the Domestication of Cultivated Tomatoes

T. Clint Nesbitt[1] and Steven D. Tanksley[2]

Department of Plant Breeding and Department of Plant Biology, Cornell University, Ithaca, New York 14850

Manuscript received December 5, 2001
Accepted for publication June 6, 2002

ABSTRACT

Sequence variation was sampled in cultivated and related wild forms of tomato at *fw2.2*—a fruit weight QTL key to the evolution of domesticated tomatoes. Variation at *fw2.2* was contrasted with variation at four other loci not involved in fruit weight determination. Several conclusions could be reached: (1) Fruit weight variation attributable to *fw2.2* is not caused by variation in the FW2.2 protein sequence; more likely, it is due to transcriptional variation associated with one or more of eight nucleotide changes unique to the promoter of large-fruit alleles; (2) *fw2.2* and loci not involved in fruit weight have not evolved at distinguishably different rates in cultivated and wild tomatoes, despite the fact that *fw2.2* was likely a target of selection during domestication; (3) molecular-clock-based estimates suggest that the large-fruit allele of *fw2.2*, now fixed in most cultivated tomatoes, arose in tomato germplasm long before domestication; (4) extant accessions of *L. esculentum* var. *cerasiforme*, the subspecies thought to be the most likely wild ancestor of domesticated tomatoes, appear to be an admixture of wild and cultivated tomatoes rather than a transitional step from wild to domesticated tomatoes; and (5) despite the fact that *cerasiforme* accessions are polymorphic for large- and small-fruit alleles at *fw2.2*, no significant association was detected between fruit size and *fw2.2* genotypes in the subspecies—as tested by association genetic studies in the relatively small sample studied—suggesting the role of other fruit weight QTL in fruit weight variation in *cerasiforme*.

DOMESTICATION of crops was one of the most profound and rapid events in plant evolution, irreversibly altering the distribution of plant species on the earth and enabling human civilization to come into existence. Domestication of individual plant species was usually enabled by one or more dramatic changes in the anatomy of the species, allowing certain desirable parts of the plant (from a human perspective) to become greatly exaggerated (*e.g.*, seed-bearing cob in maize or fruit of tomato, melon, etc.). Over recent years, evidence has accumulated to support the hypothesis that the majority of these dramatic anatomical changes can be attributed to a few loci and that selection for these loci by our ancestors rendered alterations in overall genomic diversity of the species (Doebley *et al.* 1997; Grandillo *et al.* 1999).

In 1997, Doebley *et al.* reported the cloning of *teosinte branched1* (*tb1*), a key gene associated with the evolution of wild Mexican grass teosinte into modern maize. Further studies have documented the changes in genetic variability in and around the *tb1* locus (Wang *et al.* 1999). Other than in maize, the molecular events accompanying domestication are largely unknown. Recently, however, *fw2.2*, a major quantitative trait locus (QTL) underlying the domestication of tomato, was cloned (Frary *et al.* 2000). *fw2.2* encodes a protein controlling fruit growth and mutations at this locus resulted in a major increase in fruit size during tomato domestication (Alpert *et al.* 1995; Frary *et al.* 2000). This locus makes the largest contribution to the difference in fruit size between most cultivated tomatoes and their small-fruited wild species counterparts (Alpert *et al.* 1995).

Lycopersicon (Mill.), the genus that includes the cultivated tomato, is composed of nine small-fruited species, most of which are limited in distribution to a small area in western Peru, Chile, and Ecuador (Rick 1976). Only *Lycopersicum esculentum* var. *esculentum*, the domesticated tomato, and *L. esculentum* var. *cerasiforme*, its small-fruited feral putative congener, are found outside this narrow range, being common throughout many parts of the world, especially in Mesoamerica and the Caribbean (Rick 1976). Historical and linguistic studies suggest that the cultivated tomato was most likely selected from wild forms of cerasiforme (Jenkins 1948; Rick 1976); however, phylogenetic/diversity studies based on isozymes and DNA polymorphism have not clarified this issue (Rick *et al.* 1974; Rick and Fobes 1975; Miller and Tanksley 1990; Williams and St. Clair 1993).

While the geo-historical events underlying tomato do-

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY097061–AY097189.

[1]*Present address:* Cornell Cooperative Extension, Department of Communication, Cornell University, Ithaca, NY 14850.

[2]*Corresponding author:* Department of Plant Breeding, 240 Emerson Hall, Cornell University, Ithaca, NY 14850. E-mail: sdt4@cornell.edu

TABLE 1

Lycopersicon accessions used in this study, showing the loci sequenced from each

Accession	Origin	*fw2.2*	*orf44*	*Adh2*	TG10	TG11	TG91	TG167
		L. esculentum var. *esculentum*						
M82		+	+	+	+	+	−	−
TA496		+	+	+	+	+	+	+
TA1210	"Stuffer"	+	+	+	+	+	−	−
TA1496	"Zach's giant"	+	+	+	+	+	−	−
		L. esculentum var. *cerasiforme*						
LA292	Galapagos, Ecuador	+*	−	+	+	+	+	+
LA1204	Quetzaltenango, Guatemala	+*	−	+	+	+	+	+
LA1205	Copan, Honduras	+*	−	−	−	−	−	−
LA1206	Copan, Honduras	+*	−	−	−	−	−	−
LA1226	Morona-Santiago, Ecuador	+	+	+	+	+	+	+
LA1228	Morona-Santiago, Ecuador	+*	−	−	−	−	−	−
LA1231	Napo, Ecuador	+*	−	−	−	−	−	−
LA1268	Lima, Peru	+*	−	−	−	−	−	−
LA1286	Junin, Peru	+*	−	−	−	−	−	−
LA1307	Ayachuco, Peru	+*	−	−	−	−	−	−
LA1312	Cusco, Peru	+*	−	+	+	+	+	+
LA1323	Cusco, Peru	+*	−	−	−	−	−	−
LA1324	Cusco, Peru	+*	−	−	−	−	−	−
LA1334	Arequipa, Ecuador	+*	−	−	−	−	−	−
LA1372	Lima, Peru	+*	−	−	−	−	−	−
LA1388	Junin, Peru	+*	−	+	+	+	+	+
LA1420	Napo, Ecuador	+*	−	+	+	+	+	+
LA1425	Cauca, Colombia	+*	−	−	−	−	−	−
LA1429	Manabi, Ecuador	+*	−	−	−	−	−	−
LA1455	Nuevo Leon, Mexico	+	+	+	+	+	+	+
LA1542	Turrialba, Costa Rica	+*	−	−	−	−	−	−
LA1545	Campeche, Mexico	+*	−	−	−	−	−	−
LA1549	Junin, Peru	+*	−	−	−	−	−	−
LA1574	Lima, Peru	+*	−	+	+	+	+	+
LA1619	Junin, Peru	+*	−	−	−	−	−	−
LA1621	Bahia, Brazil	+*	−	−	−	−	−	−
LA1632	Madre de Dios, Peru	+*	−	−	−	−	−	−
LA1711	Zamorano, Honduras	+*	−	−	−	−	−	−
LA1712	Pejibaye, Costa Rica	+*	−	+	+	+	+	+
LA2095	Loja, Ecuador	+*	−	−	−	−	−	−
LA2131	Zamorano-Chinchipe, Ecuador	+*	−	−	−	−	−	−
LA2616	Huanuco, Peru	+*	−	−	−	−	−	−
LA2619	Ucayali, Peru	+*	−	−	−	−	−	−
LA2664	Puno, Peru	+*	−	−	−	−	−	−
LA2675	Puno, Peru	+*	−	−	−	−	−	−
LA2688	Madre de Dios, Peru	+*	−	+	+	+	+	+
LA2845	San Martin, Peru	+*	−	−	−	−	−	−
LA2871	Sud Yungas, Bolivia	+*	−	−	−	−	−	−
LA3652	Apurimac, Peru	+*	−	−	−	−	−	−
		L. cheesmanii						
LA483	Galapagos, Ecuador	+	+	+	+	+	−	−
		L. pimpinellifolium						
LA369	Lima, Peru	+	+	+	+	+	+	+
LA1589	La Libertad, Peru	+	+	+	+	+	+	+
LA1601	Ancash, Peru	+	+	+	+	+	+	+

(*continued*)

mestication are poorly understood, even less is known about the impacts of domestication on genome diversity in tomato. Currently, *fw2.2* is the only cloned locus known to be involved in the domestication of tomato fruit. The goal of this study was to apply phylogenetic and population genetic techniques to determine the nature and origin of the mutations in *fw2.2* that have enabled domestication and to understand the impact

TABLE 1

(Continued)

Accession	Origin	*fw2.2*	*orf44*	*Adh2*	TG10	TG11	TG91	TG167
		L. parviflorum						
LA2133	Azuay, Ecuador	+	+	+	+	+	−	−
		L. peruvianum						
LA1708	Cajamarca, Peru	+	+	+	+	+	−	−
		L. hirsutum						
LA1777	Ancash, Peru	+	+	+	+	+	−	−
		L. pennellii						
LA716	Arequipa, Peru	+	+	+	+	+	−	−

+ denotes locus accessions that were sequenced in this study. +* indicates accessions from which only a 951-nt subset of the 2.7-kb 5′ UTR region of the *fw2.2* locus was sequenced (see text for details). − denotes locus-accession combinations that were not sequenced.

of domestication-related selection at the locus on the tomato genome. In an attempt to shed light on these issues, a series of *fw2.2* alleles (both coding and upstream regions) were sequenced in accessions of (1) modern tomato, (2) *L. esculentum* var. *cerasiforme*, and (3) *L. pimpinellifolium*. Variation at *fw2.2* was then contrasted with variation in other loci believed not to be involved in fruit size control: *orf44*, an anonymous gene adjacent to *fw2.2*; *Adh2* (encoding alcohol dehydrogenase); and two random, single-copy sequences, TG10 and TG11. The latter three loci are on different chromosomes than *fw2.2* and hence would not be subjected to "hitchhiking" effects due to linkage disequilibrium. These studies also permit an estimate of radiation time for the genus Lycopersicon and the divergence of cultivated tomato from its closest living wild relative, *L. pimpinellifolium*.

MATERIALS AND METHODS

Plant materials: The plant accessions used in this study are listed in Table 1. The accessions of *L. cheesmanii*, *L. hirsutum*, *L. parviflorum*, *L. pennellii*, *L. peruvianum*, and *L. pimpinellifolium* chosen for this study have been used in previous mapping populations and are known to carry alleles at the *fw2.2* locus associated with a small-fruited phenotype, referred to as "small-fruit alleles" (Grandillo *et al.* 1999). The modern cultivars of *L. esculentum* var. *esculentum* used in the study carry the "large-fruit allele" of *fw2.2* (Grandillo *et al.* 1999; S. D. Tanksley, unpublished data). Accessions of *L. esculentum* var. *cerasiforme* represent the "core collection" of the Tomato Genetic Resources Center, University of California at Davis.

Locus selection and primer design: In addition to the coding sequence (dubbed "*orfx*" in Frary *et. al.* 2000) and ~2.7 kb upstream of the *fw2.2* locus (Figure 1A), several additional loci were selected to be used as controls for sequence comparisons: (1) *orf44*, the open reading frame of unknown function immediately adjacent to *fw2.2* (Frary *et al.* 2000; see Figure 1A); (2) a 489-nucleotide region of the *Adh2* gene, including parts of exons 1–4 and introns 1–3 (Figure 1B); and (3) two unlinked single-copy genomic clones, TG10 and TG11 (Bernatzky and Tanksley 1986). The *Adh2* gene was chosen because (1) it is in on a different chromosome than *fw2.2*, (2) is a relatively highly conserved gene, containing several introns and exons in a short region, and (3) its function is not directly related to early floral organ development (Longhurst *et al.* 1994) and thus is not necessarily subject to the same selection pressures or history experienced by *fw2.2*. TG10 and TG11 are anonymous genomic sequences unlinked to *fw2.2* (chromosomes 9 and 10, respectively; Bernatzky and Tanksley 1986). The sequences of TG10 and TG11 contain no continuous open reading frames, have no significant similarity to any sequences in the GenBank nucleotide databases (BLASTN and TBLASTX), and are used to represent intragenic, relatively less-conserved noncoding sequence. For some accessions, two restriction fragment length polymorphism (RFLP) markers, TG91 and TG167, flanking the *fw2.2* region (Frary *et al.* 2000), were also sequenced. For each locus, primers were designed from available *L. esculentum* var. *esculentum* sequence, and these primer sets successfully amplified single bands in all other taxa (see below for conditions). A summary of primer sequences used for amplification is listed in Table 2.

DNA isolation, PCR amplification, purification, and sequencing: Tomato genomic DNA used for sequence analysis in this study was isolated from greenhouse-grown plants using the protocol described by Fulton *et al.* (1995). Using this DNA, PCR fragments were amplified and directly sequenced. Each PCR reaction used 0.5 μl (~100 ng) of tomato DNA and was amplified with the following thermocycler conditions: 94° denaturization (1 min), 50° annealing (1 min), and 68° elongation (2 min), for 35 cycles. PCR products used as templates for sequencing were first examined by gel electrophoresis and then cleaned using QIAGEN's (Valencia, CA) Qia-Quick spin columns. Fragments were sequenced in both directions from the same primers used for amplification, unless stated otherwise. All new sequences generated in this study have been submitted to the GenBank sequence database (accession nos. AY097061—AY097189).

Sequence analysis tools: Examination and manipulation of nucleotide sequences were conducted using the suite of programs in DNASTAR's (Madison, WI) Lasergene software package. Sequence alignments were first generated using the Clustal V method of DNASTAR Megalign (gap penalty = 10, gap length penalty = 10) and then refined by hand. Multiple sequence reads for very long regions [*fw2.2* 5′ untranslated region (UTR)] were assembled into contigs using the Phred/Phrap (Ewing and Green 1998; Ewing *et al.* 1998) and Consed (Gordon *et al.* 1998) software packages. Phylogenetic in-

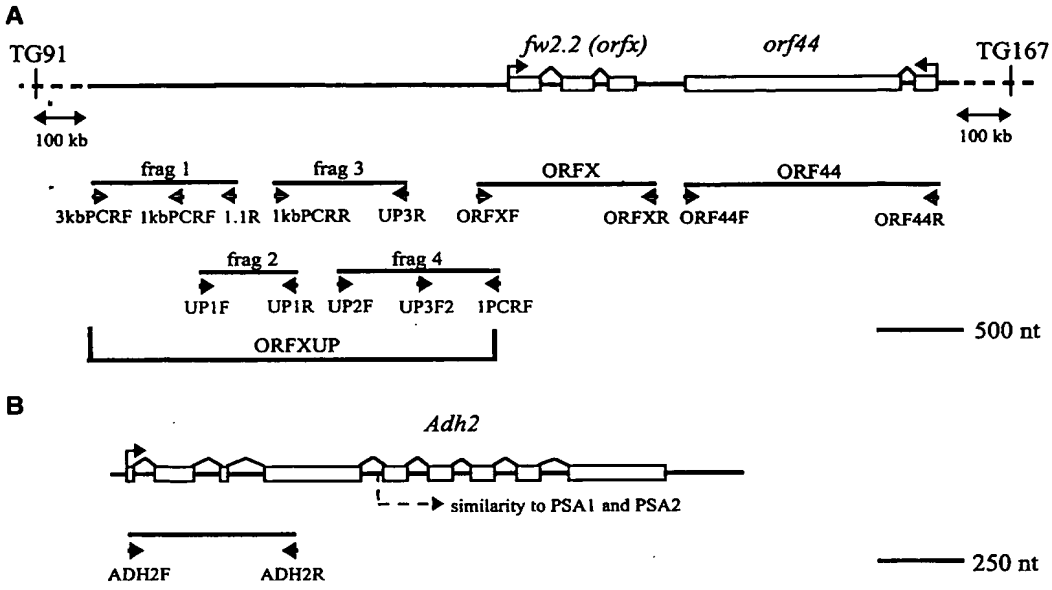


FIGURE 1.—Fragments amplified for sequence analysis. (A) The *fw2.2* region of tomato chromosome 2 (illustration based on FRARY *et al.* 2000), including the upstream and coding regions of *fw2.2* and the coding region of *orf44*. The region upstream of the *fw2.2* open reading frame was amplified as four separate fragments (frags 1–4). The positions of RFLP markers TG91 and TG167 (amplified and sequenced in some accessions) are also noted. (B) The *Adh2* gene (illustration based on LONGHURST *et al.* 1994). Amplification includes introns 1–3 and parts of exons 1–4. Note that the fragment amplified from *Adh2* does not include the region similar to *Adh2* pseudogenes PSA1 and PSA2 (see LONGHURST *et al.* 1994). Nucleotide sequences of individual primers, depicted in the figure as short arrows below each amplified fragment, are given in Table 2. Not shown: TG10 and TG11 fragments amplified and sequenced.

ferences were drawn with the assistance of β versions of PAUP* 4.0 (SWOFFORD 1998). Trees presented in this study were identified as the single most-parsimonious tree (unless stated otherwise) using a branch-and-bound search, treating gaps in the alignment as missing, and using sequence from *L. pennellii* LA716 as the outgroup. Sequence divergence estimates and other molecular population genetics statistics were generated using the DnaSP v3.53 software package (ROZAS and ROZAS 1999). Sliding-window analysis of nucleotide variability was conducted using the SWAN program of PROUTSKY and HOLMES (1998). "Statistical parsimony" analysis (TEMPLETON *et al.* 1992) used the TCS v1.13 software package (CLEMENT *et al.* 2000), and subsequent nested analysis of variance (NANOVA) used SPSS for Windows v10.0.

Fruit weight evaluation of *L. esculentum* var. *cerasiforme* accessions: To evaluate the association of fruit weight with *fw2.2* alleles among *L. esculentum* var. *cerasiforme* accessions, a single plant of each *cerasiforme* accession listed in Table 1 was grown in the field in Ithaca, New York, during the summer season of 2000. Fifteen red fruits of each accession were collected at maturity and weighed individually.

RESULTS

Sequence divergence within the genus Lycopersicon: On the basis of the sequences of the four loci examined, divergence estimates of various Lycopersicon alleles from *L. esculentum* var. *esculentum* alleles are presented in Table 3: K_s is calculated as the number of synonymous nucleotide substitutions per site, K_a is the number of nonsynonymous substitutions per site, and K is the number of substitutions per site in noncoding sequence. The values are calculated using the Jukes-Cantor method ($\alpha = 1, \beta = 1$) and represent divergence from the allele of *L. esc.* var. *esculentum* cv M82 (the allelic sequences of this accession are identical to those of other *L. esc.* var. *esculentum* accessions examined, with the exception of a single-nucleotide substitution observed in the TG10 allele of TA1210; see Figure 2). Standard errors for the divergence estimates were calculated using the method proposed by KIMURA (1980). In general, sequence divergence between species represents a few substitutions per hundred sites, even between the most distantly related species in the genus. At a few loci (*e.g.*, *Adh2*), no sequence variation was detected among some of the species tested.

To pool data from multiple loci, the significance of the variability in divergence values must be evaluated. The allelic divergence values estimated for given species pairs appear to be highly variable across loci examined. For example, K estimates for the divergence of alleles of *L. hirsutum* and *L. esculentum* cv. M82 range from ~5 to 76 substitutions per thousand sites, depending upon the locus examined. Some of this variability is likely to be due to differences in lengths of sequence examined at each locus (*i.e.*, sampling error). To test whether the observed heterogeneity is significant, a simple analysis of variance of the divergence estimates (nonzero values only) was conducted for each species comparison, using the standard errors in Table 3. In most cases, analysis of variance of K_s values could not be conducted due to the invariant nature of the sequences (*i.e.*, no variance estimates). Where analysis could be conducted on K_s estimates (*L. pimp.* LA369, *L. hirs.*, and *L. penn.*), no significant difference was found among the values. On the other hand, in most cases heterogeneity among K estimates *was* significant—*i.e.*, between-locus variation

TABLE 2

Oligonucleotide primers used for fragment amplification and sequencing

Fragment name	Primer name	Sequence	Fragment size (nt)
ORFX	ORFXF	AAATTGATGTTTTTCACCCGTTA	889
	ORFXR	ACAGGGAGTCGGAGATAGCA	
ORFXUP[a]			
Frag 1	3-kb PCRF	TGTTTAAAACGGGTCGGGTA	979
	1.1R	TTTGTTCTCTTTTCCACCGTGT	
	1-kb PCRF[b]	AGGGATACGAACAAGGAGCA	
Frag 2	UP1F	CGGTGGTCTGGACAAAATG	566
	UP1R	AACTTTATTTTAGAAAACGAAGCAAG	
Frag 3	1-kb PCRR	GGTGGTGTTGATGTGGAGTG	877
	UP3R	AAAACAAAAGGTTTAATTTACTGTCC	
Frag 4	UP2F	GATTGCGCATTGAGATGCT	934
	1PCRF	CGGGGGCAGATACATAGTGA	
	UP3F2[b]	TGAATAGGACAGTAAATTAAACCTTTT	
ORF44	44R3	CCATGAGACATGCACAAACACC	1400
	44F1	CGTGCAGGTACAAGTACAACGAATC	
ADH2	ADH2F	ATGTCGACTACTGTAGGCCAAGTC	489
	ADH2R	TCCCCTGTAAAGACAGGAAGAA	
TG10	TG10F	ATGATATCCACACCCCTGGA	587
	TG10R	ATGCCTCGAAATTCAAATGC	
TG11	TG11F	CGCGAAGATTAACCAAGAGC	586
	TG11R	TTGGGAGGCTAGATGAGGTG	
TG91	TG91F	ACGTAGGATCGGATTCGAAGT	360
	TG91R	ATCCGATCGATTAGCAGGAAT	
TG167	TG167F	ATTGCGGACTAGGCATGCATAG	520
	TG167R	GCTAGCTGGCTAACCCATGCA	

[a] To sequence the entire 2.7-kb *fw2.2* (orfx) promoter region, the region was amplified in four overlapping fragments, each sequenced separately. See Figure 1 for details.

[b] Additional internal primer used for sequencing but not for amplification.

was significantly greater than within-locus variation ($P < 0.05$). The only exception was among the K estimates between M82 and *L. cheesmanii*, which were not significantly variable. Thus, because of this significant heterogeneity among divergence estimates, any inferences based upon pooled silent-site sequence data should be made with caution. Finally, K_a values are also significantly heterogeneous among the loci (*i.e.*, in general, *orf44* is more conserved than *fw2.2*), but this result is not surprising as it is not uncommon for different genes to experience different degrees of conservation.

Estimated divergence times for the genus Lycopersicon: To provide a temporal context in which to evaluate the evolution of *fw2.2* alleles, an attempt was made to date the divergence times of species in the genus Lycopersicon. However, this exercise was done with the knowledge that rates of nucleotide substitution are notoriously variable in plants, making it extremely difficult to arrive at a suitable rate for use with molecular clock models (Muse 2000). Gaut (1998) estimated a rate of 6.03×10^{-9} synonymous substitutions per site per year (d_s) for plant nuclear genes, and a recent report applied this estimate to comparisons of *L. esculentum* and *Arabidopsis thaliana* (Ku *et al.* 2000). Given the significant locus-dependent variability in allelic divergence estimates, inferences of divergence time of species within the genus based upon this data are somewhat tenuous. Nonetheless, divergence times inferred from pooled silent-site divergence could be taken to represent very general estimates of the timing of genus radiation. Using these assumptions, Table 3 shows the estimated time, in millions of years before present (BP), that a given accession and *L. esculentum* cv. M82 diverged from a common ancestor. These results suggest that the genus Lycopersicon began its initial radiation >7 million years ago and that *L. esculentum* and its nearest relatives, *L. cheesmanii* and *L. pimpinellifolium*, diverged from a common ancestor ~1 million years BP. These dates are consistent with a recent study, which suggested that the genus Solanum, the paraphyletic taxon that includes Lycopersicon, diverged from its nearest related genus ~12 million years BP (Wikstrom *et al.* 2001).

Gene trees of Lycopersicon sequences: To evaluate the relationships among the species in the genus Lycopersicon, parsimony-based gene trees inferred from each of the sequences used in this study are shown in Figure 2. Because they introduce a large number of incongruities into the gene trees, the *cerasiforme* alleles are omitted

TABLE 3

Sequence divergence (K_s, K_a, and K) for selected nuclear loci in Lycopersicon, with standard errors

Sequence	*L. pimpinellifolium* LA369	*L. pimpinellifolium* LA1589	*L. pimpinellifolium* LA1601	*L. cheesmanii*	*L. parviflorum*	*L. peruvianum*	*L. hirsutum*	*L. pennellii*
Coding (K_s)								
fw2.2	6.3 ± 6.4	6.3 ± 6.4	6.3 ± 6.4	12.9 ± 9.2	39.1 ± 16.7	19.2 ± 11.3	26.0 ± 13.3	25.7 ± 13.2
orf44	2.1 ± 2.1	0.0	0.0	0.0	0.0	0.0	16.0 ± 6.1	10.3 ± 4.7
Adh2	0.0	0.0	0.0	0.0	0.0	0.0	18.2 ± 18.2	0.0
Coding (K_a)								
fw2.2	2.5 ± 2.5	2.5 ± 2.6	2.5 ± 2.5	3.1 ± 3.1	2.5 ± 2.5	2.5 ± 2.5	2.5 ± 2.5	5.6 ± 4.0
orf44	0.8 ± 0.8	0.8 ± 0.8	1.6 ± 1.2	0.8 ± 0.8	0.8 ± 0.8	2.5 ± 1.4	6.1 ± 2.3	6.9 ± 2.4
Adh2	0.0	0.0	0.0	0.0	0.0	0.0	9.9 ± 8.9	0.0
Noncoding (K)								
fw2.2 5′ UTR	7.8 ± 1.8	8.7 ± 2.0	8.6 ± 1.9	10.4 ± 2.1	35.8 ± 4.1	61.5 ± 5.5	76.2 ± 6.2	62.6 ± 5.6
fw2.2 introns	40.6 ± 16.9	13.3 ± 9.6	13.3 ± 9.6	6.6 ± 6.6	13.3 ± 9.6	20.1 ± 11.9	27.0 ± 13.9	55.0 ± 20.3
orf44 intron	3.2 ± 2.3	4.9 ± 2.8	3.2 ± 2.8	0.0 ± 5.9	9.8 ± 4.0	18.0 ± 5.4	21.3 ± 5.9	3.0 ± 6.2
Adh2 introns	28.2 ± 14.6	28.2 ± 10.9	28.2 ± 10.9	0.0 ± 14.2	28.2 ± 10.9	57.3 ± 15.8	48.9 ± 14.6	57.3 ± 15.8
TG10	7.0 ± 3.5	8.7 ± 3.9	7.0 ± 3.5	3.5 ± 2.5	11.2 ± 4.3	21.1 ± 6.2	8.7 ± 3.9	22.9 ± 6.4
TG11	5.2 ± 3.0	3.4 ± 2.4	5.2 ± 2.3	6.9 ± 3.5	24.4 ± 6.6	26.1 ± 6.8	5.2 ± 2.3	15.6 ± 5.2
Pooled silent sites	8.4 ± 1.3	8.0 ± 1.3	7.6 ± 1.2	6.6 ± 1.1	25.8 ± 2.3	39.7 ± 2.8	44.0 ± 3.0	41.9 ± 2.9
Divergence date	1.4 ± 0.2	1.3 ± 0.2	1.3 ± 0.2	1.1 ± 0.2	4.3 ± 0.4	6.6 ± 0.5	7.3 ± 0.5	7.0 ± 0.5

Values shown are substitutions *per thousand* sites and represent divergence from *L. esculentum* var. *esculentum* cv. M82. Also shown is an estimate of date of divergence, in million years before present, the basis of pooled silent sites and a substitution rate of 6.03×10^{-9} silent substitutions per year (from Muse 2000).

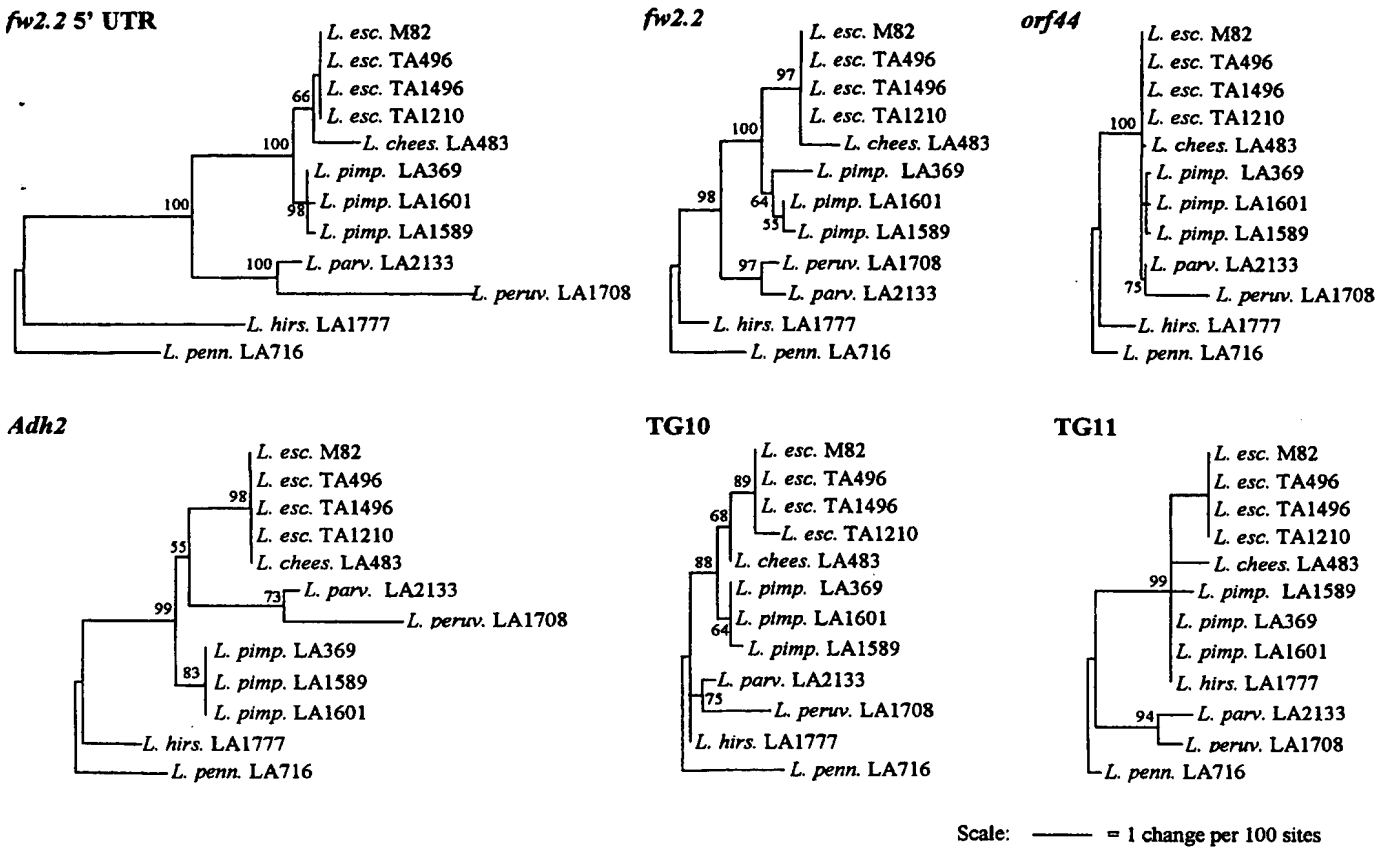


FIGURE 2.—Six gene trees of sequences from the genus Lycopersicon. *Adh2*, TG10, and TG11 are unlinked loci. The sequences used to infer trees of *fw2.2*, *orf44*, and *Adh2* include both introns and exons. To depict all trees on comparable scales, branch lengths (number of inferred steps) are divided by the length of the sequence (in nucleotides) used to construct the tree. In each case, a single most-parsimonious tree was identified. Percentages of 100 bootstrap replications are given for nodes with bootstrap values >50%. Tree statistics are as follows: *fw2.2* 5′ UTR, tree length (l) = 384, consistency index (CI) = 0.94, retention index (RI) = 0.91; *fw2.2*, l = 48, CI = 0.96, RI = 0.95; *orf44*, l = 37, CI = 1.00, RI = 1.00; *Adh2*, l = 24, CI = 0.95, LI = 0.89; TG10, l = 24, CI = 1.00, RI = 1.00; and TG11, l = 22, CI = 0.91, RI = 0.89.

from these trees for clarity and are discussed further below. In the cases of *fw2.2*, *orf44*, and *Adh2*, both introns and exons together were used to generate the trees. In general, ~500 nucleotides that include some noncoding sequence were adequate to resolve the relationships among the alleles of most species. Additionally, Figure 3 shows a tree based upon combined data.

The branching patterns of these individual and combined gene trees are generally consistent with most other published trees of the genus Lycopersicon (PALMER and ZAMIR 1982; MILLER and TANKSLEY 1990; BRETO *et al.* 1993). However, an anomalous placement of *L. hirsutum* near *L. pimpinellifolium* accessions in the TG11 tree was noted, suggesting that some lineage sorting or introgression may be associated with this species. Additionally, some sources have suggested that *L. peruvianum* may be an artificial, heterogeneous taxon (RICK 1963, 1986; MILLER and TANKSLEY 1990), having one subgroup of individuals most closely related to *L. pennellii* and *L. hirsutum* and a second group more closely related to *L. parviflorum*. The *L. peruvianum* accession used in this study, LA1708, appears to fall into the latter group.

Relative rate test: Differences in the relative rates of nucleotide substitution between lineages could be indicative of differences in past selection pressure experienced by each lineage. Selection during the process of tomato domestication could conceivably have led to a greater accumulation of nucleotide change either in the species *L. esculentum* in general or at the *fw2.2* locus in particular. To test these hypotheses, the simplified relative rate tests proposed by TAJIMA (1993) were applied to each of the five loci used in this study. Using *L. pennellii* as the outgroup, each locus was tested to determine if the *L. esculentum* var. *esculentum* sequence had evolved at a different rate than that of the sequence from *L. pimpinellifolium* or *L. cheesmanii*, its nearest wild relatives. The null hypothesis predicts that the branch length from *L. pennellii* to *L. esculentum* will be the same as the lengths from *L. pennellii* to *L. pimpinellifolium* or to *L. cheesmanii*.

For all five loci examined, using both Tajima's *D*1

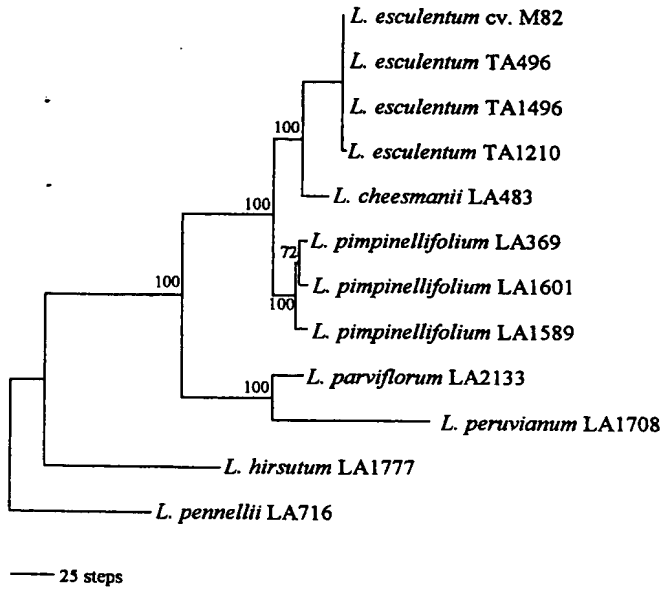


FIGURE 3.—Tree from combined sequences for the genus Lycopersicon. The sequence used to generate the phylogeny is a concatenation of all sequences used to generate trees in Figure 2 (*fw2.2* 5′ UTR, *fw2.2*, *orf44*, *Adh2*, TG10, and TG11). Percentages of 100 bootstrap replications are given for nodes with bootstrap values >50%. Tree shown is single most-parsimonious tree, length = 563, consistency index = 0.94, retention index = 0.89.

(assumes rates of transition and transversion are equal) and *D*2 (does not assume equal rates) tests, none of the test statistics were significant, providing no support for differences in mutation rates in the lineages leading to these four species. However, the statistical power of the relative rate tests is probably not very strong due to the limited number of substitutions among taxa. To increase testing power, the Tajima *D*1 and *D*2 tests were also conducted on the pooled sites from all five loci, but the test statistics were also not statistically significant in this case. Thus, neither *fw2.2* nor other tested loci appear to have diverged at a faster rate in the lineage leading to cultivated *L. esculentum*. The corollary is that there is no evidence that the *fw2.2* allele of *L. esculentum* var. *esculentum* has accumulated more (or fewer) changes than the alleles carried by related wild species.

Sequence-based inferences of functional differences between *fw2.2* alleles: Sequence analysis of the *fw2.2* region has important implications for identifying the genetic polymorphism(s) in *fw2.2* that is causally related to the variation in fruit weight associated with this locus. FRARY *et al.* (2000) reported three nonsynonymous substitutions between *L. esculentum* and *L. pennellii* in the coding region of *fw2.2*. However, further sequencing of the *fw2.2* transcription unit in other species of the genus reveals that two of the three substitutions are autapomorphies of *L. pennellii*. The third substitution (AA 3) is shared by all species of the genus except *L. esculentum* and *L. cheesmanii*; as this accession of *L. cheesmanii* is known to carry a small-fruit allele (PATERSON *et al.* 1991), this substitution is not likely to be associated with a change in fruit size. Aside from these three changes, all of the *fw2.2* alleles among the taxa examined are identical at the protein level. Furthermore, these three substitutions fall between the putative first (M1) and second (M12) methionine. Sequence-based promoter analysis, such as PROSCAN (PRESTRIDGE 1995) and the Hamming clustering method (MILANESI *et al.* 1996), fail to identify standard initiation motifs (TATA, CAAT box, CG box, etc.; reviewed in BUCHER 1990) in the vicinity of either start site. Because some uncertainty is associated with the determination of the start site, the actual start site may be M12, making all of the potentially nonsynonymous substitutions among the alleles fall in the upstream, noncoding region. In either case, the phenotypic differences between large and small alleles of *fw2.2* cannot be attributed to any functional differences in the FW2.2 protein itself.

Within the 2.7-kb region upstream of the *fw2.2* start site, only eight synapomorphies are unique to the *L. esculentum* var. *esculentum* alleles: three transitions, one transversion, and four indels 1, 2, 9, and 10 nucleotides (nt) in length, all deletions in var. *esculentum*. This suggests that the phenotype of *fw2.2* is likely to be due to one or more nucleotide changes in the upstream promoter region of the gene and supports the hypothesis that phenotypic differences may be due to differential expression of large- and small-fruit alleles (FRARY *et al.* 2000).

Sliding-window analysis (SWAN) of nucleotide variability: A sliding-window analysis was used to quantify the genus-wide nucleotide variability in the upstream UTR of *fw2.2* in an attempt to determine whether any of the eight large-fruit synapomorphies described above fall within a relatively conserved domain of the *fw2.2* promoter region. Nucleotide variability at the *fw2.2* locus (including *fw2.2* 5′ UTR, *fw2.2*, and *orf44*) was calculated using the *SWAN* software package (PROUTSKY and HOLMES 1998), and the results are shown in Figure 4. Figure 4A depicts the mean and standard deviation (SD) of nucleotide variability on the basis of the entire length of the sequence. To prevent the relatively conserved coding regions in the sequence (right half of graph) from biasing the mean and SD, Figure 4B shows the same graph, but calculates mean and SD upstream and downstream of the *fw2.2* start site separately.

In Figure 4, A and B, there are clearly regions that are conserved more highly than others, in particular the coding regions of *fw2.2* and *orf44*. Additionally, at least two regions in the *fw2.2* 5′ UTR show relatively low variability, although these "valleys" are not statistically significant (<2 standard deviations from the mean in both graphs). None of the eight large-fruit synapomorphies in the promoter region of *fw2.2* (marked with "Δ") appear to fall within well-conserved regions—on

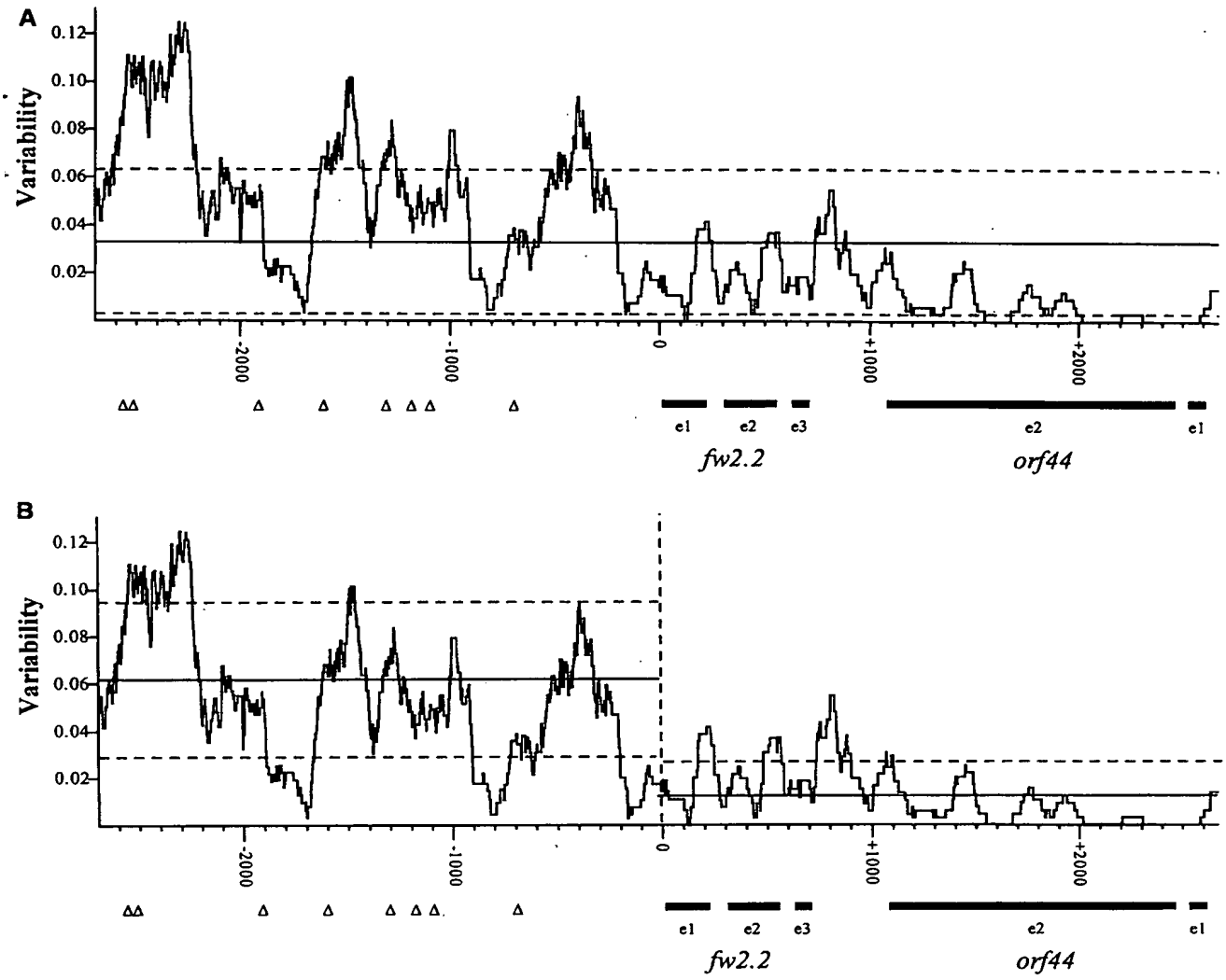


FIGURE 4.—SWAN of nucleotide variability in the *fw2.2* region, including the 5′ UTR region of *fw2.2*, the *fw2.2* transcription unit, and the adjacent *orf44* transcription unit (depicted with heavy bars below the graph). (A) Mean and standard deviation of variability calculated from the full length of the sequence. (B) Mean and standard deviation calculated separately for the regions upstream (left) and downstream (right) of the *fw2.2* start site. Solid horizontal bar, mean variability; dashed horizontal bars, one standard deviation from the mean. The positions of the eight large-fruit synapomorphies are denoted "Δ" beneath the graph. Nucleotide position numbers are relative to the *fw2.2* start site. Sequences used for the calculation include all non-*esculentum* accessions shown in Figure 2 (eight accessions total). Accessions of *L. esculentum* were omitted from the SWAN analysis to prevent putative "large allele" mutations from adding to the calculation of variability.

the contrary, they seem to lie in areas of average or higher variability. If any of the eight large-fruit synapomorphies do in fact fall within an important, conserved domain, those domains may be so short as to not stand out against the background of random variation in sequence variability along the length of the alignment.

Diversity of *L. esculentum* var. *cerasiforme* alleles across five loci: Because small-fruited *L. esculentum* var. *cerasiforme* is thought to be the wild progenitor of the large-fruited domesticated cultivars, a 951-nucleotide fragment of the *fw2.2* 5′ UTR (spanning five of the eight large-fruit synapomorphies) was sequenced from a sample of 39 *cerasiforme* accessions. The coding region of *fw2.2* was not examined among the *cerasiformes*, as previous results suggested polymorphisms in this region are not likely to be important to variation in fruit size. The allelic diversity among the *cerasiforme* accessions, with sequences of the same fragment from the *L. esculentum* var. *esculentum*, *L. cheesmanii*, and *L. pimpinellifolium* accessions examined above, is depicted by the gene tree in Figure 5. Seven different haplotypes were identified among the *cerasiforme* accessions (denoted A–G). Most of the *cerasiforme* accessions carry the haplotype identical to the domesticated, large-fruited *esculentum* varieties.

Figure 5 also includes the country of origin of the accessions examined. Although the B haplotype—the

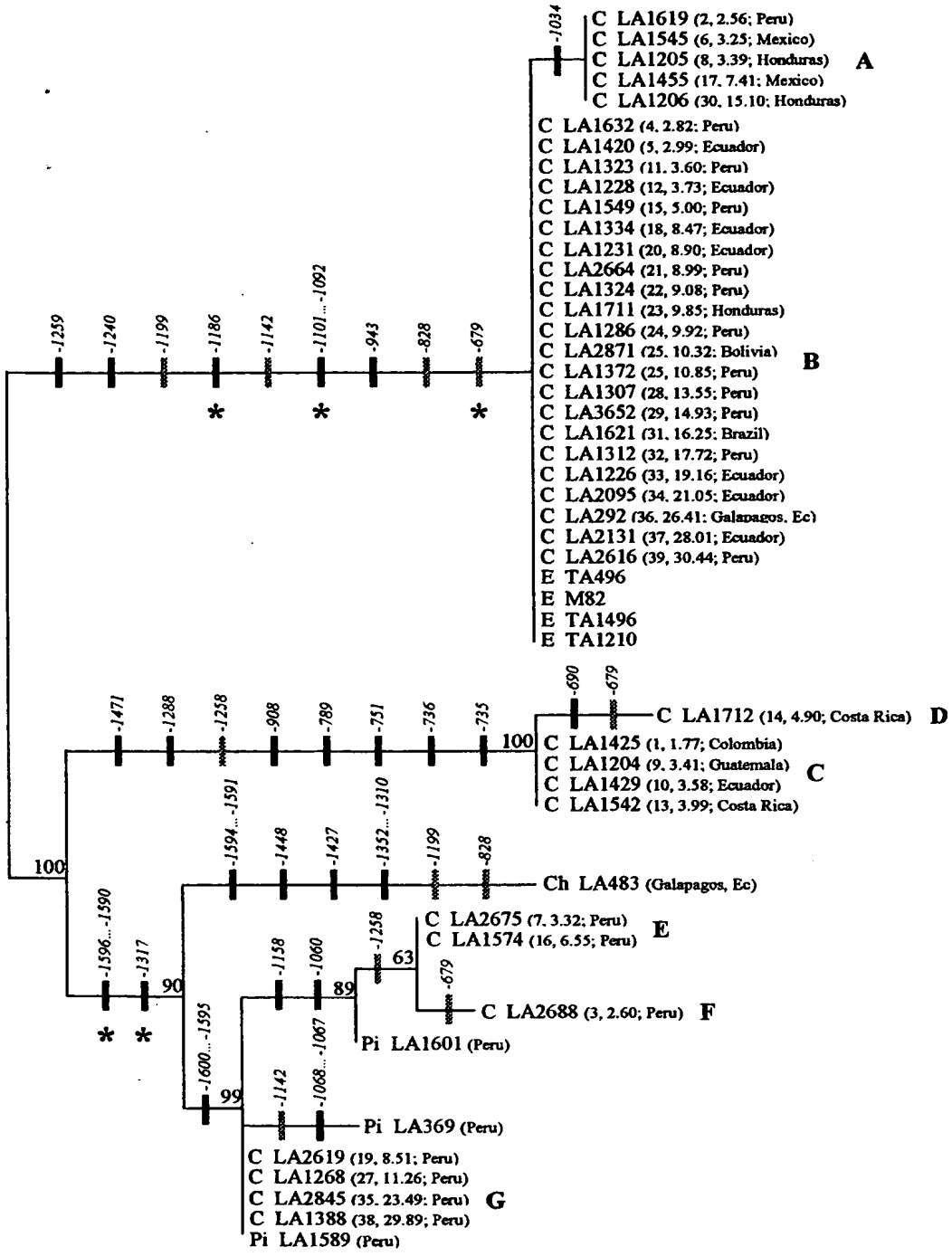


FIGURE 5.—Gene tree of sequences from *L. pimpinellifolium* (Pi), *L. cheesmanii* (Ch), *L. esculentum* var. *esculentum* (E), and *L. esculentum* var. *cerasiforme* (C), based on a 951-nucleotide subset of the *fw2.2* 5′ UTR. Tree shown is the single most-parsimonious tree, using deletions as a fifth state (tree length = 79, consistency index = 0.9241, retention index = 0.9250). Percentages of 100 bootstrap replications are given for nodes with bootstrap values >50%. The placements of character changes on the tree are as in the most-parsimonious tree; vertical hatch marks on branches denote individual substitutions or indels inferred along each branch, numbered by alignment position upstream of the *fw2.2* start site. Solid hatches denote synapomorphies, and shaded hatches denote inferred homoplasies. Of the eight large-fruit allele synapomorphies in the *fw2.2* 5′ UTR (discussed in text), five are included in this tree and are marked with asterisks. The seven haplotypes observed among the *cerasiforme* accessions are denoted with boldface letters (A–G) to the right of the tree. Also included, in parentheses after the *cerasiforme* accession numbers, are (1) the overall rank in mean fruit weight among the 39 *cerasiforme* accessions (with 1 = smallest weight), (2) mean fruit weight (in grams) of each accession, and (3) the country of origin.

allele identical to the "large allele" carried by var. *esculentum*—is distributed throughout the natural geographical range of var. *cerasiforme*, haplotypes E, F, and G appear to be restricted in distribution to areas sympatric with *L. pimpinellifolium* (Peru). Haplotypes A, C, and D are also found in areas sympatric with *L. pimpinellifolium*, in Ecuador and Peru, but are more frequently found outside this region.

To contrast allelic diversity of *fw2.2* with the rest of the genome, *Adh2*, TG10, and TG11 sequences from a sample of 10 of the 39 *cerasiformes* were examined. *Cerasiforme* alleles at each locus appear as a paraphyletic clade with members grouping with alleles either from the domesticated *esculentum* or from the *L. pimpinellifolium* accessions (Figure 6). Moreover, *cerasiforme* alleles fall into different subclades, depending on which gene is examined. LA292 (C3 in Figure 6), for example, carries an *esculentum*-like allele at *fw2.2* and *Adh2*, but a

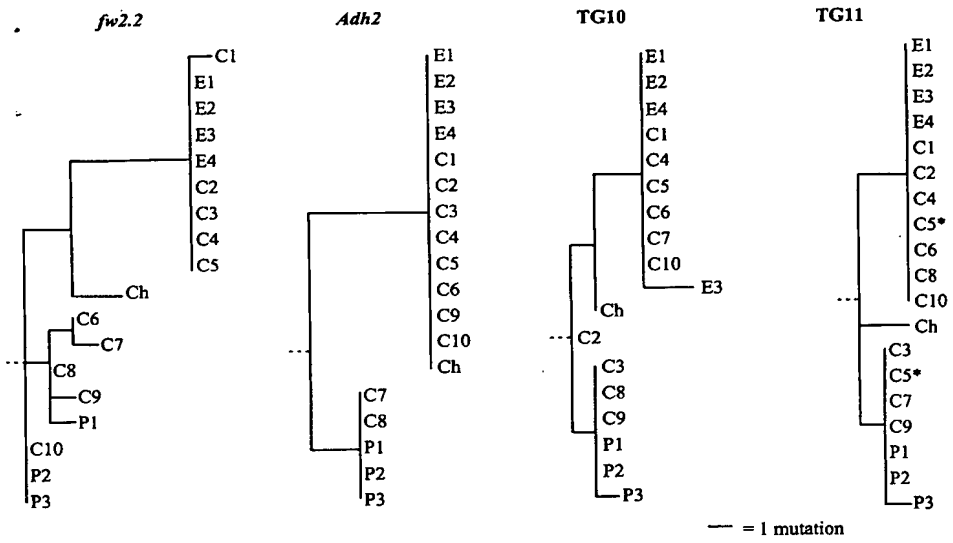


FIGURE 6.—Allelic diversity at four loci within a sample of 10 *cerasiformes*. The four trees correspond to four of the six gene trees in Figure 2 (*fw2.25'* UTR, *Adh2*, TG10, TG11) and show branching only among accessions of *L. esculentum* var. *esculentum* (E), *L. esculentum* var. *cerasiforme* (C), *L. cheesmani* (Ch), and *L. pimpinellifolium* (P). Accessions included are: M82 (E1), TA496 (E2), TA1496 (E3), TA1210 (E4), LA1455 (C1), LA1226 (C2), LA292 (C3), LA1312 (C4), LA1420 (C5), LA1204 (C6), LA1712 (C7), LA1574 (C8), LA2688 (C9), LA1388 (C10), LA483 (Ch), LA1601 (P1), LA369 (P2), and LA1589 (P3). The asterisks in the TG11 tree denote that accession C5 (LA1420) appears to be heterozygous at this locus. All trees are drawn to the same scale, representing the number of inferred steps (for clarity, *not* corrected for sample length as in Figure 2). Sequence data from *fw2.2* and *orf44* were not collected for this subset of taxa.

pimpinellifolium-like allele at TG10 and TG11. In contrast, the small set of domesticated *esculentums* always group together. In fact, with the exception of a single-nucleotide difference in the TG10 allele of TA1496 (E3), no allelic diversity was observed among the *esculentums*. The *cerasiformes* thus represent a diverse population containing an admixture of both *esculentum*- and *pimpinellifolium*-like alleles and suggest that the subspecies may be derived from hybridizations between *L. esculentum* domesticates and *L. pimpinellifolium* wild forms.

If the presence of *pimpinellifolium*-like alleles represents recent introgression into *L. esculentum* var. *cerasiforme* from *L. pimpinellifolium*, then some linkage disequilibrium may be detectable by observing closely linked markers. TG91 and TG167, two RFLP markers flanking the *fw2.2* region by <0.1 cM or 100 kb upstream and downstream, respectively (see Figure 1; FRARY *et al.* 2000), were also sequenced in the accessions used above. Although there were polymorphisms between the *L. esculentum* var. *esculentum* and *L. pimpinellifolium* alleles at both loci, the 10 *L. esculentum* var. *cerasiforme* accessions were monomorphic and identical to the *L. esculentum* var. *esculentum* allele at both loci. Because the same accessions were polymorphic for *fw2.2* alleles, this suggests that if the *pimpinellifolium*-like alleles are introgressions from *L. pimpinellifolium*, they must have occurred far enough in the past that linkage to TG91 and TG167 has been broken. TG91 and TG167 are also the only markers observed in this study for which no *pimpinellifolium*-like alleles are detected among the *cerasiformes* (with the caveat that only 10 accessions were sampled).

Molecular population genetics analysis of *L. esculentum* accessions: Sequence-based genetic analysis was performed on *L. esculentum* accessions (both *cerasiforme* and cultivated types) to make inferences about the history of *L. esculentum* population structure. A summary of basic population statistics is presented in Table 4. The most striking result in the table is the near absence of polymorphism among the four modern cultivars—only a single-nucleotide substitution in one var. *esculentum* accession was observed in a sample of >7 kb. While the sample of cultivars is small, it contained a sample of diverse types. Two accessions (M82 and TA496) are modern processing tomatoes producing "roma-type" fruit and two (TA1210 and TA1496) are heirloom varieties, one with extremely large fruit (TA1496) and one with bell-pepper-shaped fruit (TA1210). This lack of variation in var. *esculentum* is consistent with previous surveys of var. *esculentum* diversity, which determined levels of polymorphism among cultivated tomatoes to be *extremely* low (MILLER and TANKSLEY 1990). This lack of diversity is most likely a reflection of at least three population bottlenecks in the history of modern cultivars: (1) initial domestication, (2) transfer of varieties to Europe by Spanish explorers, and (3) subsequent breeding efforts by primarily U.S. breeders (RICK 1976).

Many population models infer historic selection pressures on the basis of observed violations of neutral nucleotide substitutions (KIMURA 1980). For example, on the basis of this neutral theory, the Hudson-Kreitman-Aguadé (HKA) test predicts that loci that evolve at higher rates should have higher levels of within-species

TABLE 4

Estimates of nucleotide diversity within the species *L. esculentum*

		L. esc. var. *esculentum*				*L. esc.* var. *cerasiforme*				*L. esculentum* overall			
Locus	Sites	*n*	π	θ	*D*	*n*	π	θ	*D*	*n*	π	θ	*D*
fw2.2 5′ UTR	951	4	0	0	—	39	0.0054	0.0041	1.03	43	0.0050	0.0040	0.83
Adh2	498	4	0	0	—	10	0.0055	0.0054	0.02	14	0.0041	0.0048	−0.60
TG10	592	4	0.0017	0.0019	−0.71	10	0.0035	0.0024	1.77	14	0.0033	0.0032	0.15
TG11	586	4	0	0	—	10	0.0026	0.0018	1.68	14	0.0022	0.0016	1.08

For each of the four loci examined, summary data listed include number of nucleotide sites sampled at the locus (sites), the number of accessions examined within each taxon (n), nucleotide diversity (π), estimated θ values per site (θ; from η, number of mutations), and Tajima's D test statistic (TAJIMA 1993). None of the D values is significant ($\alpha = 0.05$), and three values could not be calculated (—) due to absence of substitutions.

polymorphism as compared to polymorphism at other neutral loci (HUDSON *et al.* 1987). Similarly, MCDONALD and KREITMAN (1991) proposed the test that for neutral loci the ratio of fixed nonsynonymous to synonymous substitutions between species should be equal to the same ratio within species. However, the absence of polymorphism within var. *esculentum* accessions sequenced in this study limits the ability to apply these neutral theory-based methods [*i.e.*, HKA tests whether 0 = 0; McDonald-Kreitman (MK) causes division by 0 errors]. That nucleotide diversity among var. *esculentum* alleles appears to be lower than diversity among var. *cerasiforme* alleles is consistent with a population bottleneck in the history of var. *esculentum*. However, nucleotide polymorphism at the level of individual genes may not be adequate to make robust population inferences about past selection pressures without a potentially prohibitively large amount of nucleotide sequence—>7 kb from many more than four accessions.

Test for association of genotype and fruit weight phenotype in *L. esculentum* var. *cerasiforme*: Mean fruit weight (from a 15-fruit sample) of each of the 39 *cerasiforme* accessions studied was superimposed upon the gene tree in Figure 5. The phenotypic data provide an ideal opportunity to evaluate so-called "measured genotypes" (BOERWINKLE *et al.* 1987)—in this case, to assign fruit-weight effects to individual haplotypes of the *fw2.2* 5′ UTR. Clearly there is a large range in fruit weight among the *cerasiformes*—a nearly 12-fold difference from smallest to largest. Due to sequence identity or similarity to alleles of known phenotype (*i.e.*, the alleles carried by the var. *esculentum*, *L. pimpinellifolium*, and *L. cheesmanii* accessions), the initial expectation was that plants carrying haplotypes A and B would have significantly larger fruit than those carrying all other alleles (C–G). Yet, although the *cerasiformes* in the A–B clade have slightly larger fruit (mean = 10.3 g, SD = 7.8) than those in the C–G clade (mean = 7.4 g, SD = 8.2), this difference is not significant (one-tailed *t*-test, $P = 0.146$).

To attribute phenotypic effects to individual haplotypes, the NANOVA method proposed by TEMPLETON *et al.* (1987) was utilized. This method is based upon the assumption that changes in phenotype follow the same evolutionary history represented by the cladogram and is therefore dependent upon (1) confidence in the cladogram and (2) the assumption that recombinant alleles are rare. The Templeton-Crandall-Sing (TCS) methods (TEMPLETON *et al.* 1992; CLEMENT *et al.* 2000) were used to evaluate these assumptions. First, all seven *cerasiforme* haplotypes can be assembled into a single network (with no closed loops, which would signify recombination) within the 95% parsimony limit (13 steps)—*i.e.*, each step within the cladogram is likely to be parsimonious. Second, although the cladogram contains a number of homoplasies, no recombinant alleles could be identified using the TCS method; in particular, there were no postulated recombination events that could resolve *two or more* homoplasies (AQUADRO *et al.* 1986). Therefore, NANOVA was performed using the most parsimonious tree of *cerasiforme* haplotypes.

The nesting categories used for NANOVA are illustrated in Figure 7. Because many of the haplotypes are separated by multiple steps—requiring a large number of inferred, intermediate haplotypes that make no statistical contribution to the model—a modification of the grouping method of TEMPLETON *et al.* (1987) was used. Rather than strictly nest the groupings on the basis of single-step increments, nesting categories were based more generally upon "subclades." The lowest level of nesting (level 0) represents individual haplotypes, labeled A–G. The next level of nesting (level 1) groups the haplotypes into four subclades, haplotypes A and B (1), C and D (2), E and F (3), and G (4), and the highest nesting level (level 2) divides the taxa into two groups, A and B (I) and C–G (II). Thus, the NANOVA model of fruit weight variance contains three terms: variation among level 2 clades, variation among level 1 clades within level 2 clades, and variation among level 0 clades within level 1 clades within level 2 clades.

The results of NANOVA are summarized in Table 5. As with the one-tailed *t*-test above, the contrast expected

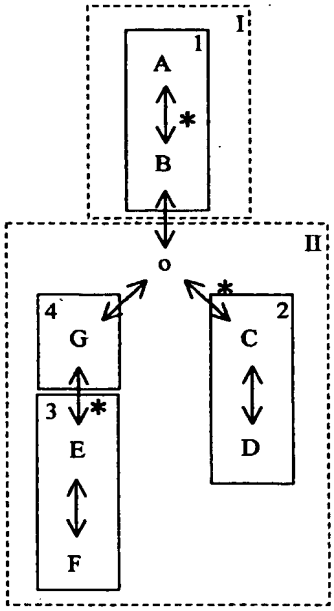


FIGURE 7.—Haplotype categories used for nested analysis of variance, following the procedure outlined by TEMPLETON *et al.* (1987). A–G represent the seven *cerasiforme* haplotypes of the *fw2.2* 5′ UTR, as defined in Figure 5— *i.e.*, the "level 0 clades." Arrows depict the phylogenetic relationships among the seven haplotypes inferred in Figure 3 (note: arrows represent multiple steps and are not drawn to scale). Solid lines enclose the four "level 1 clades" designated by numbers (1–4), and dashed lines enclose the two "level 2 clades" designated by Roman numerals (I and II). The small circle represents an inferred intermediate haplotype, and its exact categorical placement is irrelevant to the statistical analysis. Asterisks indicate those branches inferred to be significantly associated with variation in fruit weight.

to be most significant—variation between level 2 clades—is not significant. That is, there is no evidence that the fruit of plants carrying putative "large alleles" (inferred from sequence identity) are significantly larger than those carrying putative "small alleles." However, several other terms in the model are significant. First, there is significant variation among level 1 clades. All of this variation can be attributed to variation among level 1 clades *within level 2 clade II,* because there is only one level 1 clade within level 2 clade I. Multiple comparisons among the three level 1 clades in level 2 clade II [using the Bonferroni method to account for multiple comparisons (NETER *et al.* 1985)] reveals that the contrasts of clades 2 *vs.* 4 and 3 *vs.* 4 are significant ($P = 0.015$ and 0.038, respectively), but 2 *vs.* 3 is not. Finally, the only significant variation identified within level 1 clades (that is, among haplotypes) was within level 1 clade 1; variation between haplotypes A and B is significant. Thus, the NANOVA method has identified three branches in the cladogram that have a change in fruit size associated with them: (1) the branch between haplotypes A and B, (2) the branch between level 1 clades 3 and 4, and (3) the branch between level 1 clade 2 and its inferred common ancestor with clade 4. These significant branches are marked with asterisks in Figure 7.

Lack of significance might suggest that the mutations associated with the *fw2.2* phenotype may fall outside the sequenced promoter region and are not in perfect linkage disequilibrium with that region. Or, perhaps more likely, a large portion of the fruit weight variation in *cerasiforme* may be attributable to polymorphism at several of the other known fruit weight quantitative trait loci (GRANDILLO *et al.* 1999), and the contribution of *fw2.2* is too small to be detected against this background. Further, it should be noted that although significant associations were detected between phenotype and some branches in the cladogram, it is not necessarily true that mutations along those branches *cause* the observed phenotype. Rather, phenotype could also be caused by changes outside of the sequenced region that are in linkage disequilibrium with those observed mutations. Finally, it is curious to note that the haplotypes

TABLE 5

Nested analysis of variance of fruit weight among 39 *L. esculentum* var. *cerasiforme* accessions, following the method of TEMPLETON *et al.* (1987)

Source	Type III sum of squares	d.f.	Mean square	*F*-statistic	Significance
Level 2 clades (total)	18.13	1	18.13	0.32	0.58
Level 1 clades (total)	491.19	2	245.60	4.32	0.02
Within II	491.19	2	245.60	4.32	0.02
Level 0 clades (total)	177.67	3	59.22	1.04	0.39
Within 1	171.69	1	171.69	3.02	0.09
Within 2	2.35	1	2.35	0.04	0.84
Within 3	3.64	1	3.64	0.06	0.80
Error	1817.61	32	56.80		

The accessions are nested by the *fw2.2* 5′ UTR sequence into seven distinct haplotypes (level 0 clades), as shown in Figure 5. These seven haplotypes in turn are nested into four level 1 clades, which in turn are nested into two level 2 clades, as illustrated in Figure 7.

most significantly associated with decreased fruit size (A, C, and D) are observed in accessions outside the natural range of *L. pimpinellifolium*.

DISCUSSION

The *fw2.2* phenotype cannot be explained by differences in protein structure or function. Instead, data presented here support the observation of FRARY *et al.* (2000) that the fruit-size phenotype is likely due to differences in expression of the gene, probably as a result of one of eight mutations in the 2.7 kb upstream of the *fw2.2* gene. Further, the very low rate of nonsynonymous substitutions among the coding sequences of most taxa examined here (*fw2.2*, *orf44*, and *Adh2*) suggests that much of the phenotypic diversity within the genus may be due to the changes within the noncoding sequences in the genome. Although the observation of considerably lower diversity among var. *esculentum* accessions relative to var. *cerasiforme* accessions is consistent with a bottleneck in the history of var. *esculentum*, genetic diversity among var. *esculentum* accessions is too low to make neutral theory-based inferences about historic selection pressures. That is, the distinction between a selective sweep and neutral lineage sorting cannot be made at the loci examined. Tajima's relative rate test, however, does suggest that the large-fruited var. *esculentum* allele of *fw2.2* has not accumulated more (or fewer) substitutions than other alleles in the genus.

Phylogenies of Lycopersicon have been inferred using a variety of molecular methods: chloroplast DNA (PALMER and ZAMIR 1982), mitochondrial DNA (MCCLEAN and HANSON 1986), RFLPs (MILLER and TANKSLEY 1990), and isozymes (BRETO *et al.* 1993). This study represents the first reconstruction of Lycopersicon phylogeny based upon the sequence of individual nuclear loci. Although sequence distances between species are not great, they are generally large enough to produce robust phylogenies from a sample of 300–500 nucleotides. However, among *L. pimpinellifolium*, *L. esculentum* var. *cerasiforme*, and *L. esculentum* var. *esculentum*, incongruities are observed (Figures 5 and 6), which may be due to the fact that these species are entirely interfertile and gene flow among them has been well documented in areas where they are sympatric (RICK 1950, 1958; RICK *et al.* 1974; RICK and FOBES 1975; RICK and HOLLE 1990; WILLIAMS and ST. CLAIR 1993). Frequent introgressions among these taxa make it extremely difficult, if not impossible, to track the exact origins of individual alleles—var. *cerasiforme* appears to represent an admixture of alleles from *L. esculentum* varieties and *L. pimpinellifolium*. There are *pimpinellifolium*-like alleles among *cerasiforme* accessions collected in areas that are not sympatric with *L. pimpinellifolium* (some accessions with haplotypes C and D). Although there is certainly a great deal of gene flow within *L. esculentum*, it also seems unlikely that the high proportion of large-fruit alleles among the *cerasiformes* could be explained entirely by recent introgressions from domesticated types. Thus is it probably reasonable to infer that the allelic diversity among the *cerasiformes* today is not entirely a result of recent introgression and may be similar to the diversity that would have been available to early tomato cultivators.

Because fruit of the *cerasiformes* are already considerably larger than those of the other members of the genus (RICK 1958), it is conceivable that the large allele of *fw2.2* arose not among relatively recent domesticates selected from the *cerasiformes* in Mesoamerica, but further in the past, perhaps predomestication, when *L. esculentum* var. *cerasiforme* first diverged from the other species in the genus in the Andes. If these molecular-clock-based divergence dates are reasonably accurate (Table 3), then the large and small alleles could have diverged from a common ancestor >1 million years BP. Although the conversion of an *fw2.2* allele from small phenotype to large need only have been the most recent substitution in its divergence from its common ancestor with *L. pimpinellifolium*, *fw2.2* may have acquired its large-fruit nature long before humans entered the New World (WENKE 1990).

Unlike *teosinte branched1*, *fw2.2* is a QTL and does not condition a dramatic morphological change in tomato fruit, but rather an *incremental* one. An association of large-fruit phenotype with presence of putative large-fruit alleles of *fw2.2* could not be detected among *cerasiformes* accessions against the background of what are likely to be many other genes affecting fruit weight in tomato (GRANDILLO *et al.* 1999). The range in fruit size among the *cerasiforme* accessions examined here is >15 times greater than the difference in size between near-isogenic lines differing at the *fw2.2* locus (ALPERT *et al.* 1995). If the variation in *cerasiforme* fruit size present today is at all representative of the variation present for the early agriculturalists, then they might not have even noticed a spontaneous mutation in the *fw2.2* locus. Instead, the evolution of fruit size during the domestication of tomato is likely to represent a very long path of lineage sorting and gene "stacking" of alleles at many loci—some of which, including the large allele of *fw2.2*, could have existed for millennia before the first Americans.

The authors thank J. Doyle, M. Jahn, and A. Clark for their critical review of this manuscript and L. Patton and A. Patton for their encouragement in the field. The work was supported by grants from the U.S. Department of Agriculture National Research Initiative Cooperative Grants Program (no. 96-35300-36460), the Binational Research and Development Fund (no. US 2427-94), and the National Science Foundation (no. 9872617) to S.D.T.

LITERATURE CITED

ALPERT, K. A., S. GRANDILLO and S. D. TANKSLEY, 1995 *fw2.2*: a major QTL controlling fruit weight is common to both red- and green-fruited tomato species. Theor. Appl. Genet. **91**: 994–1000.

AQUADRO, C. F., S. F. DEESE, M. M. BLAND, C. H. LANGLEY and C. C.

LAURIE-AHLBERG, 1986 Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. Genetics **114:** 1165–1190.

BERNATZKY, R., and S. D. TANKSLEY, 1986 Toward a saturated linkage map in tomato based on isozymes and cDNA sequences. Genetics **112:** 887–898.

BOERWINKLE, E. S., S. VISVIKIS, D. WELSH, J. STEINMETZ, S. M. HANASH *et al.*, 1987 The use of measured genotype information in the analysis of quantitative phenotypes in man. II. The role of apolipoprotein E polymorphism in determining levels, variability and covariability of cholesterol, betalipoprotein and triglycerides in a sample of unrelated individuals. Am. J. Med. Genet. **27:** 567–582.

BRETO, M. P., M. J. ASINS and E. A. CARBOMELL, 1993 Genetic variability in Lycopersicon species and their genetic relationships. Theor. Appl. Genet. **86:** 113–120.

BUCHER, P., 1990 Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoters. J. Mol. Biol. **212:** 563–578.

CLEMENT, M., D. POSADA and K. A. CRANDALL, 2000 TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9:** 1657–1659.

DOEBLEY, J., A. STEC and L. HUBBARD, 1997 The evolution of apical dominance in maize. Nature **386:** 485–488.

EWING, B., and P. GREEN, 1998 Basecalling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:** 186–194.

EWING, B., L. HILLIER, M. WENDL and P. GREEN, 1998 Basecalling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:** 175–185.

FRARY, A., T. C. NESBITT, A. FRARY, S. GRANDILLO, E. VAN DER KNAAP *et al.*, 2000 *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. Science **289:** 85–88.

FULTON, T. M., J. CHUNWONGSE and S. D. TANKSLEY, 1995 Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol. Biol. Rep. **13** (3): 207–209.

GAUT, B. S., 1998 Molecular clocks and nucleotide substitution rates in plants. Evol. Biol. **30:** 93–120.

GORDON, D., C. ABAJIAN and P. GREEN, 1998 Consed: a graphical tool for sequence finishing. Genome Res. **8:** 195–202.

GRANDILLO, S., H. M. KU and S. D. TANKSLEY, 1999 Identifying the loci responsible for natural variation in fruit size and shape in tomato. Theor. Appl. Genet. **99:** 978–987.

HUDSON, R., M. KREITMAN and M. AGUADÉ, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

JENKINS, J. A., 1948 The origin of the cultivated tomato. Econ. Bot. **2:** 379–392.

KIMURA, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

KU, H. M., T. VISION, J. LIU and S. D. TANKSLEY, 2000 Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. Proc. Natl. Acad. Sci. USA **97:** 9121–9126.

LONGHURST, T., E. LEE, R. HINDE, C. BRADY and J. SPEIRS, 1994 Structure of the *Adh2* gene and *Adh2* pseudogenes, and a study of the *Adh2* gene expression in fruit. Plant Mol. Biol. **26:** 1073–1084.

MCCLEAN, P. E., and M. R. HANSON, 1986 Mitochondrial DNA sequence divergence among *Lycopersicon* and related *Solanum* species. Genetics **112:** 649–667.

MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *adh* locus in *Drosophila*. Nature **351:** 652–654.

MILANESI, L., M. MUSELLI and P. ARRIGO, 1996 Hamming clustering method for signal prediction in 5′ and 3′ regions of eukaryotic genes. Comput. Appl. Biosci. **12** (5): 399–404.

MILLER, J. C., and S. D. TANKSLEY, 1990 RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor. Appl. Genet. **80:** 437–448.

MUSE, S. V., 2000 Examining rates and patterns of nucleotide substitution in plants. Plant Mol. Biol. **42:** 25–43.

NETER, J., W. WASSERMAN and M. K. KUTNER, 1985 *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*, Ed. 2. Richard D. Irwin, Homewood, IL.

PALMER, J. D., and D. ZAMIR, 1982 Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon*. Proc. Natl. Acad. Sci. USA **79:** 5006–5010.

PATERSON, A. H., S. DAMON, J. D. HEWITT, D. ZAMIR, H. D. RABINOVICH *et al.*, 1991 Mendelian factors underlying quantitative traits in tomato: comparisons across species, generations, and environments. Genetics **127:** 181–197.

PRESTRIDGE, D. S., 1995 Predicting Pol II promoter sequences using transcription factor binding sites. J. Mol. Biol. **249:** 923–932.

PROUTSKY, V., and E. C. HOLMES, 1998 SWAN: sliding window analysis of nucleotide sequence variability. Bioinformatics **14:** 467–468.

RICK, C. M., 1950 Pollination relations of *Lycopersicon esculentum* in native and foreign regions. Evolution **4:** 110–122.

RICK, C. M., 1958 The role of natural hybridization in the derivation of cultivated tomatoes of western South America. Econ. Bot. **12:** 346–367.

RICK, C. M., 1963 Barriers to inbreeding in *Lycopersicon peruvianum*. Evolution **17:** 216–232.

RICK, C. M., 1976 Tomato, pp. 268–273 in *Evolution of Crop Plants*, edited by N. W. SIMMONDS. Longman Group, London.

RICK, C. M., 1986 Reproductive isolation in the *Lycopersicon peruvianum* complex, pp. 477–495 in *Solanaceae Biology and Systematics*, edited by W. G. DARCY. Columbia University Press, New York.

RICK, C. M., and J. F. FOBES, 1975 Allozyme variation in the cultivated tomato and closely related species. Bull. Torrey Bot. Club **102:** 376–386.

RICK, C. M., and M. HOLLE, 1990 Andean *Lycopersicon esculentum* var. *cerasiforme*: genetic variation and its evolutionary significance. Econ. Bot. **44:** 69–78.

RICK, C. M., R. W. ZOBEL and J. F. FOBES, 1974 Four peroxidase loci in red-fruited tomato species: genetics and geographic distribution. Proc. Natl. Acad. Sci. USA **71:** 835–836.

ROZAS, J., and R. ROZAS, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

SWOFFORD, D. L., 1998 *PAUP* 4.0: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, MA.

TAJIMA, F., 1993 Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135:** 599–607.

TEMPLETON, A. R., E. BOERWINKLE and C. F. SING, 1987 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. Genetics **117:** 343–351.

TEMPLETON, A. R., K. A. CRANDALL and C. F. SING, 1992 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132:** 619–633.

WANG, R. L., A. STEC, J. HEY, L. LUKENS and J. DOEBLEY, 1999 The limits of selection during maize domestication. Nature **398:** 236–239.

WENKE, R. J., 1990 *Patterns in Prehistory: Humankind's First Three Million Years*. Oxford University Press, New York.

WIKSTROM, N., V. SAVOLAINEN and M. W. CHASE, 2001 Evolution of the angiosperms: calibrating the family tree. Proc. R. Soc. Lond. Ser. B **1482:** 2211–2220.

WILLIAMS, C. E., and D. A. ST. CLAIR, 1993 Phenetic relationships and levels of variability detected by restriction fragment length polymorphism and random amplified polymorphic DNA analysis of cultivated and wild accessions of *Lycopersicon esculentum*. Genome **36:** 619–630.

Communicating editor: T. F. C. MACKAY